

MASTER'S THESIS

De ontwikkeling van een statistische methode voor het evalueren van de intuïtiviteit van de PGA-modelleertaal

Schilders, C.

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

Open Universiteit
www.ou.nl



BPMIT graduation project

BPMIT graduation assignment preparation (IM0602)

Business Process Management and IT Graduation Assignment (IM9806)



De ontwikkeling van een statistische methode voor het evalueren van de intuïtiviteit van de PGA-modelleertaal

The development of a statistical method for evaluating the intuitiveness of the PGA modeling language

Opleiding:	Open Universiteit, faculteit Betawetenschappen Masteropleiding Business Process Management & IT
Programme:	Open University of the Netherlands, faculty of Science Master of Science Business Process Management & IT
Cursus:	IM0602 Voorbereiden Afstuderen BPMIT IM9806 Afstudeeropdracht Business Process Management and IT
Student:	Carron Schilders
Identiteitsnummer:	
Datum:	12-10-2021
Afstudeerbegeleider	Dhr. B. Roelens
Meelezer	Mevr. L. Cuijpers
Derde beoordelaar	n.v.t.
Versie nummer:	1.0
Status:	Definitief

Abstract

Om domeinen en hun onderlinge relaties in detail te kunnen beschrijven zijn er speciale modelleertalen ontworpen, de zgn. *domain specific modeling languages (DSMLs)*. Omdat een DSML gericht is op een specifiek domein heeft zij ook een iconografie die afgestemd is op dat domein met als doel de betrokken eindgebruikers een duidelijke visualisatie te geven waardoor er effectief geanalyseerd en gecommuniceerd kan worden. De notatie speelt daarom een belangrijke rol, deze dient voor de gebruiker herkenbaar te zijn voor het domein dat gemodelleerd wordt anders wordt hij/zij cognitief te zwaar belast en wordt het model niet of moeilijk te interpreteren. De *Process-Goal-Alignment (PGA)* modelleertaal is de DSML taal die in dit onderzoek wordt geëvalueerd op zijn intuïtiviteit. Het uitgevoerde onderzoek betreft een experiment dat *eye-tracking* en een online *questionnaire* gebruikt waarmee de verbetervoorstellen uit eerder onderzoek statistisch worden geëvalueerd. Uit de resultaten van dit onderzoek blijkt dat de elementen *customer goal*, *valueStream* en *importance* intuïtief blijken te zijn in de nieuwe notatie. Voor de laatste twee elementen komt dit door hun verbindende rol in de taal en voor het eerste element door zijn unieke grafisch ontwerp.

Sleutelbegrippen

Eye-tracking, Intuïtiviteit, Process-Goal-Alignment, strategic fit, Understandability

Samenvatting

Om domeinen en hun relaties in detail te kunnen beschrijven zijn er speciale modelleertalen ontworpen de zgn. *domain specific modeling language (DSML)*. Omdat een DSML gericht is op een specifiek domein heeft zij ook een iconografie die afgestemd is op dat domein met als doel de betrokken eindgebruikers een duidelijke visualisatie te geven waardoor er effectief geanalyseerd en gecommuniceerd kan worden over een bepaald probleem binnen zo'n domein. De notatie speelt daarom een belangrijke rol bij een modelleertaal, deze dient voor de gebruiker herkenbaar te zijn voor het domein dat gemodelleerd wordt anders wordt hij/zij cognitief te zwaar belast en wordt het model niet of moeilijk te interpreteren.

Eén van de DSML talen is de *Process-Goal-Alignment (PGA)* modelleertaal. De taal modelleert de strategische afstemming binnen een bedrijfsarchitectuur en stelt zo bedrijven in staat om te kunnen reageren op kansen en bedreigingen. De eindgebruikers zijn stakeholders die binnen een organisatie een direct belang hebben bij een strategische afstemming. Daarom was één van de belangrijke *requirements* voor het ontwerpen van deze taal, dat deze een intuïtieve notatie moest hebben. Volgens Bork et al. (2019) zorgt een intuïtieve notatie in een DSML ervoor dat de cognitieve belasting bij de eindgebruiker verschoven wordt naar een perceptuele taak waardoor de hersenen minder belast worden.

In dit onderzoek wordt de intuïtiviteit van de PGA-modelleertaal geëvalueerd waarbij we uitgaan van een eerder onderzoek van Roelens & Bork (2020). Uit dat onderzoek kwamen een aantal verbetervoorstellen naar voren die tot stand zijn gekomen door een combinatie van creativiteit en interpretatie van de onderzoekers. Doel van dit onderzoek is dan ook om de resultaten uit dat kwalitatief onderzoek statistisch te evalueren.

De hoofdvraag luidt dan als volgt: "Hoe kan de intuïtiviteit van verschillende versies van de PGA notatie statistisch worden vergeleken?"

Er wordt in dit onderzoek een experiment opgezet met een twintigtal studenten bedrijfskundige informatica die in twee groepen worden gesplitst, nl. één groep met de oude notatie en één met de nieuwe notatie. Beide groepen kregen dezelfde vragen van een *online questionnaire* met visualisaties. De vragen waren opgesplitst in twee categorieën, nl. *comprehension questions* en *problem solving tasks*. Bij de eerste categorie moesten de respondenten vragen beantwoorden over de elementen en bij de tweede categorie moesten de respondenten een probleem oplossen.

Daarnaast werden tijdens het uitvoeren van het experiment de oogbewegingen van de respondenten vastgelegd d.m.v. *eye-tracking*. Deze triangulatie van verschillende bronnen zorgt ervoor dat de variabele die relevant zijn voor de intuïtiviteit, gemeten en geïnterpreteerd kunnen worden. Deze variabelen zijn: (1) *AOI-fixatie run count*, (2) *AOI-fixatieduur*, (3) *totale fixatieduur*, (4) *totale fixatie aantal*, (5) *effectiveness*, (6) *interpretational efficiency* en (7) *perceived ease of use*.

Uit de resultaten van dit onderzoek blijkt dat de elementen *customer goal*, *valueStream* en *importance* uit de nieuwe notatie intuïtief zijn gebleken en dat *valueStream* en *importance* een belangrijke rol spelen in zowel de *comprehension questions* als *problem solving tasks* vanwege hun

verbindende en associatieve rol binnen de taal. De *customer goal* dankt zijn intuïtiviteit door zijn unieke grafisch ontwerp.

Summary

To be able to describe domains and their relationships, special modeling languages have been designed, the so-called domain specific modeling languages (DSMLs). Because a DSML is focused on a specific domain, it also has an iconography that is tailored to that domain with the aim of giving the end users involved a clear visualization so that they can effectively analyze and communicate about a certain problem within such a domain. The notation therefore plays an important role in a modeling language, it must be recognizable to the user for the domain that is being modeled, otherwise he/she will be cognitively overloaded and the model will be difficult or even impossible to interpret.

The Process-Goal-Alignment (PGA) modeling language is one of those DSML languages. The language models strategic alignment within a business architecture, enabling companies to respond to opportunities and threats. The end users are stakeholders within an organization who have a direct interest in strategic coordination. Therefore, one of the important requirements for designing this language was that it should have an intuitive notation. According to Bork et al. (2019), an intuitive notation in a DSML ensures that the cognitive load on the end user is shifted to a perceptual task, which puts less stress on the brain.

In this study the intuitiveness of the PGA modeling language is evaluated, based on a previous study by Roelens & Bork (2020). A number of suggestions for improvement emerged from that research, which were the result of a combination of the researchers' creativity and interpretation. The aim of this research is therefore to statistically evaluate the results of that qualitative research.

The main question is then: "How can the intuitiveness of different versions of the PGA notation statistically evaluated?"

In this research, an experiment is set up with approximately twenty students of business informatics who are split into two groups, i.e. one group with the old notation and one with the new notation. Both groups received the same questions from an online questionnaire with visualizations. The questions were split into two categories, i.e. 'comprehension questions' and 'problem solving tasks'. In the first category, the respondents had to answer questions about the elements and in the second category, the respondents had to solve a problem.

In addition, during the experiment, the eye movements of the respondents were recorded by means of eye tracking. This triangulation of different sources ensures that the variables relevant to the intuitiveness can be measured and interpreted. These variables are: (1) AOI fixation run count, (2) AOI fixation duration, (3) total fixation time, (4) total fixation number, (5) effectiveness, (6) interpretational efficiency and (7) perceived ease of use.

The results of this research show that the elements customer goal, valueStream and importance have been found to be intuitive from the new notation and that valueStream and importance play an important role in both the comprehension questions and problem solving tasks because of their

connecting and associative role within the language. The customer goal owes its intuitiveness to its unique graphic design.

Inhoudsopgave

Abstract.....	iii
Sleutelbegrippen.....	iii
Samenvatting.....	iv
Summary.....	v
Inhoudsopgave.....	vii
1. Introductie	1
1.1. Achtergrond	1
1.2. Gebiedsverkenning	3
1.3. Probleemstelling	4
1.4. Opdrachtformulering	4
1.5. Motivatie / relevantie	5
1.6. Aanpak in hoofdlijnen	5
2. Theoretisch kader	6
2.1 Onderzoeksaanpak.....	6
2.2. Uitvoering.....	8
2.3. Resultaten en conclusies.....	8
2.4. Doel van het vervolgonderzoek	10
3. Methodologie.....	10
3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)	10
3.2. Technisch ontwerp: uitwerking van de methode	11
3.3. Gegevensanalyse.....	14
3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten	15
4. Resultaten	17
4.1. Uitvoering.....	17
4.2. Demografische gegevens	18
4.3. Algemene resultaten.....	19
4.4. Hypotheses H-fixatie duur en H-fixatie aantal.....	20
4.5. Hypotheses H-AOI fixatie duur en H-AOI <i>run count</i>	30
4.6. Hypotheses H- <i>Comprehension questions effectiveness</i> en H- <i>Problem solving effectiveness</i>	35
4.7. H- <i>Interpretational efficiency comprehension questions</i> en H- <i>Interpretational efficiency problem solving tasks</i>	40
4.8. H- <i>Perceived ease of use</i>	43

4.9.	AOI run count en AOI fixatie duur - Score.....	45
4.10.	Responsetijd – Score.....	50
5.	Discussie, conclusies en aanbevelingen.....	52
5.1.	Discussie.....	52
5.1.1.	Resultaten	53
5.2.	Reflectie	59
5.2.1.	Interne validiteit.....	59
5.2.2.	Betrouwbaarheid	60
5.3.	Conclusies	61
5.3.1.	Deelvraag 1: Wat is intuïtie en waar bestaat deze uit?	61
5.3.2.	Deelvraag 2: Hoe kan intuïtie gemeten en gekwantificeerd worden?	61
5.3.3.	Deelvraag 3: Hoe kan de relatie gelegd worden tussen intuïtie en de visuele perceptie van conceptuele modelleertalen?	61
5.3.4.	Deelvraag 4: Hoe kunnen de gevonden resultaten uit het theoretisch kader vertaald worden in een concreet uitvoerbare methode?	62
5.3.5.	Deelvraag 5: Hoe intuïtief is de nieuw voorgestelde PGA notatie in vergelijking met de oorspronkelijke versie?.....	62
5.4.	Aanbevelingen voor de praktijk.....	63
5.5.	Aanbevelingen voor verder onderzoek.....	64
6.	Referenties.....	65
7.	Bijlage 1 Opzet experiment van Roelens & Bork (2020)	66
8.	Bijlage 2 Zoek queries voor bronnen	68
9.	Bijlage 3 Conceptuele modellen	70
10.	Bijlage 4 Literatuuranalyse.....	74
11.	Bijlage 5 Data en preparatie	86
11.1.	Algemeen (weghalen <i>outliers</i>)	86
11.2.	Sectie <i>comprehension questions</i> en <i>problem solving tasks</i> data preparatie algehele fixatie duur en aantal.....	92
11.3.	Sectie <i>comprehension questions</i> en <i>problem solving tasks</i> data preparatie AOI data	95
11.4.	Sectie <i>comprehension questions</i> en <i>problem solving tasks</i> data preparatie effectiviteit	98
11.1.	Sectie <i>comprehension questions</i> en <i>problem solving tasks</i> data preparatie efficiëntie.....	99
11.2.	Sectie <i>comprehension questions</i> en <i>problem solving tasks</i> data preparatie ease of use ..	103
11.3.	Data bestanden.....	106
12.	Bijlage 6 Effectgrootte berekening	106
13.	Bijlage 7 Communicatie aan respondenten.....	107
14.	Bijlage 8 PGA modellen.....	108

1. Introductie

1.1. Achtergrond

Het onderzoek speelt zich af op het gebied van *Enterprise Architecture* (EA). Elk van deze woorden hebben een eigen betekenis. *Enterprise* is een verzameling van organisaties met een gemeenschappelijk doel en *Architecture* of “Architectuur is de fundamentele organisatie van een systeem bestaande uit componenten en hun relaties tot elkaar en hun omgeving, geleid door haar ontwerp-principes” (Lankhorst, 2013, p.2). Wanneer de twee voorgaande termen worden gecombineerd tot EA, luidt de definitie hiervan als volgt:

- Enterprise Architecture bestaat uit een “samenhangend geheel van principes, methodes en modellen die worden gebruikt bij het ontwerpen en realiseren van een *enterprise*’ organisatie structuur, bedrijfsprocessen, informatiesystemen en infrastructuur” (Lankhorst, 2013, p.11).

EA is uit te leggen met de metafoor om een huis te bouwen. De architectuur is het ontwerp van hoe het huis gebouwd moet worden. Het ontwerp bestaat uit meer dan alleen maar plaatjes of een tekening. Zaken als constructieregels, materiaalkeuze, het volgen van bepaalde ontwerpregels, zijn terug te vinden in een ontwerp. Vakmensen die vanuit het ontwerp het huis moeten bouwen zijn in staat om vanuit hun eigen domein (loodgieters of elektriciens vak, etc.) het ontwerp te lezen en te begrijpen en van gedachten te wisselen met de architect.

Anders dan bij het ontwerpen van een huis, zijn bij het opbouwen of ontwerpen van een organisatie de stappen die genomen moeten worden niet gestandaardiseerd. Het is mogelijk om te beginnen met het ontwerpen van bedrijfsprocessen of te beginnen bij het applicatielandschap dat misschien al aanwezig was en waarop de bedrijfsprocessen moeten aansluiten. *The Open Group Architectural Framework* (TOGAF) (The Open Group, 2020) is een aantal jaren geleden als standaard ontwikkeld voor het nemen van deze stappen. Het geeft *enterprise* architecten de mogelijkheid een organisatie te ontwerpen op een uniforme en gestandaardiseerde manier. Zo’n *framework* of raamwerk zorgt voor “een coherent geheel van principes, methodes en modellen om een geïntegreerd beeld te geven van de strategie, structuur, processen, informatie en technologie domeinen van een organisatie” (Lankhorst, 2013, p.11).

Een goede EA levert inzichten in wat nodig is om (nieuwe) bedrijfsregels (denk bijv. aan AVG regels) te balanceren en door te vertalen naar strategie en dagelijkse operaties. Om dit te kunnen bereiken moeten stakeholders en betrokkenen vanuit de verschillende domeinen de architectuur kunnen begrijpen. Tot 2009 ontbrak het nog aan een uniforme en gebruikersvriendelijke modelleertaal om een *enterprise* architectuur te visualiseren. Dit gebrek werd ingevuld door ArchiMate. ArchiMate (The Open Group, 2020) is een modelleertaal die tools levert om een *enterprise* architectuur en zijn domeinen te beschrijven, te analyseren en de relaties tussen de onderlinge domeinen te visualiseren. Deze modelleertaal gebruikt generieke notaties voor bedrijfselementen. Vaak zijn dit geometrische figuren zoals rechthoeken of cirkels. Deze zijn bewust gekozen vanwege de brede insteek.

Bij het modelleren van een bepaald domein worden vaak domein specifieke modelleertalen (DSMLs) gebruikt om de concepten van zo'n domein te kunnen visualiseren gericht op de betrokken eindgebruikers, wat met een generieke notatie niet mogelijk is. De notatie van een DSML bevat bijv. iconen die de aspecten van het domein visueel weergeven. De iconen hebben tot doel de cognitieve belasting van de lezer van het model te ontlasten zodat het model beter gebruikt kan worden daar waar het voor bedoeld is, nl. voor o.a. het analyseren en het communiceren over een bepaald probleem (Roelens & Bork, 2020).

Voor een efficiënte communicatie m.b.v. een model, speelt de notatie dus een belangrijke rol. Het is immers het eerste contact dat gebruikers hebben met het model. Voor de gebruiker moet de notatie herkenbaar zijn voor het domein dat gemodelleerd is. Volgens Bork, Schruffer, and Karagiannis (2019) pleit dit er dan ook voor dat een modelleertaal (en zeker een DSML) een intuïtieve notatie taal moet zijn, waardoor er bij het lezen van modellen een verschuiving ontstaat van een cognitieve taak naar een perceptuele taak die voor de hersenen minder belastend en intensief is.

Wie of wat bepaalt nu of een modelleertaal intuïtief is? Dat is natuurlijk de eindgebruiker, degene die het model moet lezen en begrijpen. De ontwikkelaar van een modelleertaal kan misschien beargumenteren waarom er gekozen is voor een bepaald icoon om een bepaald element te visualiseren, maar als dat niet overeenkomt met de verwachtingen van de eindgebruiker dan is de keuze voor dat icoon niet intuïtief. Er zal dus afstemming moeten plaatsvinden met eindgebruikers om te bepalen of een modelleertaal als intuïtief beschouwd kan worden.

Eén van de DSML talen die binnen het terrein van EA wordt toegepast is de PGA-modelleertaal. De visualisatie van de taal is opgebouwd uit een *business architecture heatmap* die d.m.v. kleurcodes de performance en strategische waardecreatie binnen een hiërarchie van bedrijfsmiddelen weergeeft (Roelens, Steenacker, & Poels, 2019). De *business architecture heatmap* bestaat uit een achttal elementen, nl. *Customer Goal*, *Internal Goal*, *Financial Goal*, *Process*, *Activity*, *Value Proposition*, *Competence* en *Financial Structure*, die weer worden verbonden met zgn. *valueStreams* of waardestromen. De waardestromen en elementen vormen samen een hiërarchische structuur waarbij elk element in de hiërarchie zijn eigen waarde creëert of m.a.w. zijn bijdrage levert om de strategische doelstelling of afstemming van de organisatie te bereiken. Hoe belangrijk en in welke mate een element zijn waarde bijdraagt binnen de hiërarchie, wordt aangegeven door *Performance* en *Importance* indicatoren. Kort samengevat kan men stellen dat de taal erop gericht is om strategische afstemming te realiseren binnen de bedrijfsarchitectuur (Roelens & Bork, 2020). Het stelt een organisatie in staat om te kunnen reageren op kansen of bedreigingen. Volgens Roelens et al. (2019) is de eindgebruiker van de PGA techniek iedere stakeholder die binnen de organisatie belang heeft bij strategische afstemming en toegang heeft tot de hiervoor relevante informatie. Daarom was één van de ontwerp *requirements* dat business eindgebruikers de *business architecture heatmap* op een intuïtieve manier moeten kunnen begrijpen. De intuïtiviteit van de PGA taal was echter nog niet eerder getest. In de paper van Roelens & Bork (2020) wordt dan ook een methode voorgesteld, die de intuïtiviteit van de PGA modelleertaal evalueert. In dit onderzoek zal verder worden gegaan op deze evaluatiemethode.

1.2. Gebiedsverkenning

In de paper van Roelens & Bork (2020) wordt de evaluatie techniek van Bork et al. (2019) toegepast om de intuïtiviteit van de PGA-modelleertaal te testen in samenwerking met een aantal eindgebruikers. De evaluatie bestond uit een aantal taken die samengebundeld waren tot een drietal fases (zie figuur 1). Voor de gedetailleerde opzet van het experiment wordt verwezen naar [bijlage 1](#).

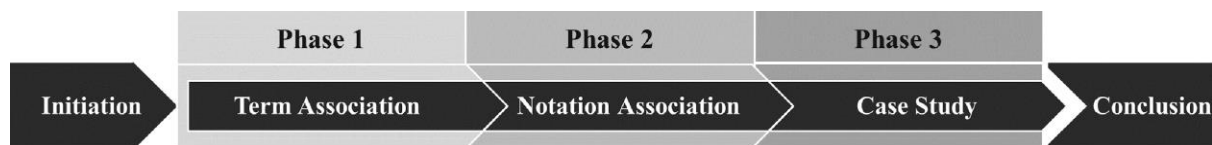


Fig. 1: Fases voorkomend in de evaluatiemethode voor de intuïtiviteit van een DSML (Bork et al., 2019)

Initiation

De initiatie bestond uit een korte introductie over het domein (i.e. strategische afstemming) en de bouwblokken van de PGA modelleertaal zonder de respondenten de eigenlijke notatie te laten zien. Verder hebben de eindgebruikers moeten aangeven wat hun geslacht en leeftijd is en wat hun ervaring is met modelleertalen. Er is ze meegedeeld dat de gecollecteerde data anoniem zal blijven.

Fase één - Term Association

In deze fase is aan de respondenten gevraagd om a.d.h.v. een lijst met termen die voorkomen binnen de PGA-modelleertaal, intuïtief een grafische voorstelling te maken.

Fase twee - Notation Association

De respondenten kregen voorbeelden van de PGA notatie te zien zonder daarbij extra informatie te verstrekken over de achterliggende betekenis. De bedoeling was om bij elke notatie een term te noteren, die bij het zien ervan werd opgeroepen.

Fase drie - Case Study

Bij de case study kregen de respondenten een *business architecture heatmap* te zien van een organisatie zonder een legenda erbij te voegen. Aan de respondenten werden begripsvragen gesteld over het model. Deze vragen kunnen beantwoord worden op basis van de informatie die in het model vervat zit.

Conclusie

Als laatste stap werd de respondenten gevraagd om kwalitatieve feedback en evt. verbetervoorstellen aan te brengen op de bestaande PGA - notatie.

Na het analyseren van de drie fases en de conclusie voor de PGA modelleertaal, zijn er in het totaal zes nieuwe notatie suggesties gedaan voor de volgende PGA elementen: *Competence*, *Value Proposition*, *Internal Goal*, *Customer Goal*, *valueStream* en *Importance*. Voor vijf elementen is besloten geen nieuwe notatie voor te stellen. Van de nieuwe notaties kan echter niet vastgesteld worden of deze werkelijk een verbetering zullen zijn. Kortom, het onderzoek levert geen statistisch bewijs of de nieuw voorgestelde notatie juist wel of niet werkt.

1.3. Probleemstelling

De voorgestelde aanpassingen aan de PGA notatie komen voort uit kwalitatief onderzoek. Dit betekent dat ze deels tot stand gekomen zijn door analyses, maar ook door creativiteit en interpretatie van de onderzoekers. Men heeft echter geen echt bewijs voor de conclusies omdat de bevindingen niet statistisch onderschreven kunnen worden. Dit onderzoek zal gericht zijn om de conclusies van het kwalitatieve onderzoek statistisch te onderbouwen.

De onderzoeksvraag luidt dus als volgt:

Hoe kan de intuïtiviteit van verschillende versies van de PGA notatie statistisch worden vergeleken?

1.4. Opdrachtformulering

Het doel van dit onderzoek is om de intuïtiviteit van verschillende versies van de PGA notatie te evalueren.

Uit het experiment van Roelens & Bork (2020) zijn een aantal verbeterde PGA notaties voorgesteld die deels door menselijke creativiteit en inspiratie tot stand zijn gekomen. Het is echter niet vastgesteld of deze nieuwe notatie daadwerkelijk intuïtiever is dan de oude. In dit onderzoek wordt met behulp van statistische methoden de intuïtiviteit van de nieuwe en oude notatie geëvalueerd.

Om te komen tot het antwoord op de hoofdvraag moeten eerst een aantal deelvragen worden beantwoord:

1. Wat is intuïtie en waar bestaat deze uit?
2. Hoe kan intuïtie gemeten en gekwantificeerd worden?
3. Hoe kan de relatie gelegd worden tussen intuïtie en de visuele perceptie van conceptuele modelleertalen?
4. Hoe kunnen de gevonden resultaten uit het theoretisch kader vertaald worden in een concreet uitvoerbare methode?
5. Hoe intuïtief is de nieuw voorgestelde PGA notatie in vergelijking met de oorspronkelijke versie?

In dit onderzoek wordt er dieper ingegaan op intuïtiviteit. De eerste drie deelvragen worden beantwoord uit het literatuuronderzoek waarna deelvraag vier wordt beantwoord met het opzetten van de evaluatiemethode zelf. De antwoorden zullen inzicht geven in welke data er verzameld moet worden, hierbij kan gedacht worden aan een online of fysiek kwantitatief interview en/of een experiment. In deelvraag vijf zullen de antwoorden afkomstig uit het experiment geanalyseerd en gekwantificeerd worden om een besluit te kunnen trekken over de mate van intuïtiviteit van beide PGA notaties. Dit besluit zal vervolgens leiden tot het antwoord op de hoofdvraag.

1.5. Motivatie / relevantie

Volgens Braatz & Brandt (2012) draagt een DSML bij aan de acceptatie van een scenario dat normaal gesproken geanalyseerd wordt a.d.h.v. tekstuele beschrijvingen en informele modelleertalen. Een DSML is ontwikkeld voor een specifiek domein waardoor deze ook alle concepten bevat van dat domein. Eindgebruikers uit zo'n domein zijn hierdoor snel in staat het model te interpreteren en evt. problemen te detecteren, waarna een gerichte discussie kan plaatsvinden tussen de betrokkenen. De snelheid wordt voor een groot deel bepaald door hoeverre de visuele elementen aansluiten op het specifieke domein. Binnen het domein van softwareontwikkeling zijn veel DSML-talen uitgerust met code generatie functionaliteiten. Hierdoor kan met een eenvoudige druk op de knop code gegenereerd worden vanaf het meta-model. Op deze manier ontstaat exact hetgeen wat is gemodelleerd, waar het *development team* verder op kan door ontwikkelen. Daarnaast verkleint het vaak de mogelijke interpretatieverschillen tussen business en IT binnen een organisatie.

Een DSML is zeer geschikt om de complexiteit te verlagen en de focus te leggen op relevante aspecten binnen een domein. Hierbij is het zeer belangrijk dat de notatie van de modelleertaal zo intuïtief mogelijk is. Een intuïtieve modelleertaal vermindert het cognitief proces en verhoogt het perceptuele proces dat veel minder intensief is. Daarnaast zorgt intuïtiviteit ervoor dat adoptie van de taal door de eindgebruikers sneller plaatsvindt. De vraag naar intuïtieve modelleertalen groeit nog steeds, alleen ontbreekt het binnen de wetenschappelijke literatuur aan een evaluatie methode waarmee de intuïtiviteit gemeten kan worden (Bork et. al. 2019). De resultaten van dit onderzoek zullen een bijdrage leveren aan een statistische evaluatie methode die toegepast kan worden op DSML talen in het algemeen.

1.6. Aanpak in hoofdlijnen

Na de de inleiding komen de volgende hoofdstukken aan bod.

In het Theoretisch Kader (Hoofdstuk 2) zal op zoek worden gegaan naar bevindingen uit de wetenschappelijke literatuur die antwoord zullen geven op de relatie tussen intuïtiviteit en de visuele perceptie van conceptuele modelleertalen.

In het hoofdstuk Methodologie wordt de manier van uitvoering van het empirisch onderzoek beschreven en hoe de bevindingen uit het Theoretisch Kader zijn verwerkt in dit onderzoek.

De resultaten van het empirisch onderzoek zullen in Hoofdstuk 4 worden beschreven waarna in Hoofdstuk 5 conclusies getrokken zullen worden uit het onderzoek en er een discussie volgt op de vraag in hoeverre de resultaten uit dit onderzoek te vergelijken zijn met het eerder onderzoek.

2. Theoretisch kader

2.1. Onderzoeksaanpak

Algemene configuratie:

Als basis is de OU bibliotheek genomen met daar in de volgende standaard configuraties opgenomen: geen datum constraints, alle content typen, alle disciplines, Engelse taal, peer reviewed publications, boek reviews werden niet meegenomen in de zoekresultaten en er werd gezocht binnen de collectie van de OU bibliotheek.

Het stop criterium ligt bij 30. Hierna werd verondersteld geen relevante papers meer te vinden

Deelvraag 1: Wat is intuïtie en waar bestaat deze uit?

Door alleen het woord 'intuïtie' in de zoekquery te gebruiken waren er teveel resultaten waarvan verreweg de meeste niet relevant waren, dus werd de zoek query aangepast naar: (*Understanding*) AND (*Intuition*) Hier kwam ook geen bevredigend resultaat uit en is de zoekquery aangepast naar: (*measure*) AND (*intuition*) AND (*Problem Solving*).

De drie zoekwoorden zijn gekozen omdat deze de kern omschrijven van wat we willen weten over intuïtie, nl. of er een manier bestaat intuïtie te meten en in hoeverre intuïtie een rol speelt bij het oplossen van een probleem. Dit leverde de volgende paper op: *Intuition, intelligence, data compression* (Kipper, 2019). De paper is weliswaar niet geciteerd, maar na het lezen van de abstract was het artikel de moeite waard om door te nemen omdat er een poging wordt gedaan om intuïtie theoretisch te omschrijven en er een link gelegd moet worden naar een *computational* model. De paper week echter teveel af van waar ik naar op zoek was waardoor de paper toch niet relevant genoeg was. Door op deze paper *backward snowballing* toe te passen kwam ik op de paper van Okoli, Weller & Watt, (2016) *Information processing and intuitive decision-making on the fireground: towards a model of expert intuition*. In deze paper viel het op dat de auteur Hogarth veel werd aangehaald. Op de paper van Okoli et al. (2016) is daarom *backward snowballing* toegepast en is uiteindelijk een relevante paper gevonden van Hogarth, nl.: *Educating Intuition: A Challenge for the 21st Century* (Hogarth, 2003). De paper beschrijft helder welke cognitieve processen en denksystemen een rol spelen in wat wij noemen, 'intuïtie'.

Deelvraag 2: Hoe kan intuïtie gemeten en gekwantificeerd worden?

Voor deelvraag 2 is er gekozen om de *peer reviewed* papers optie uit te vinken. Het bleek dat de papers met deze optie aan niet relevant bleken te zijn, dit was wel het geval wanneer deze optie werd uitgezet.

In eerste instantie zijn de woorden *intuition*, *measure* en *quantify* opgenomen in de zoekquery. Hier kwamen meer dan 10000 resultaten uit. Er is één paper, deze stond op de tweede plek, vluchtig bekeken, nl. *Superficial wall stress: the long awaited comprehensive biomechanical parameter to objectify and quantify our intuition* (Levy, Ajjawi, & Roberts, 2010), maar ontdekte snel dat de paper te vol zat met terminologiën uit de biochemie waardoor het lastig was een goed beeld te vormen. Aan bovenstaande zoekquery is *visual perception* toegevoegd om de query meer specifiek te maken.

Dit leverde één artikel op wat in de richting kwam van wat ik zocht nl. *Quality Metrics for Information Visualization* (Behrisch, Blumenschein, Kim, Shao, El-Assady, Fuchs, Seebacher, Diehl, Brandes, Pfister, Schreck, Weiskopf & Keim, 2018). De paper verwijst naar twee relevante artikelen die relevant zijn voor deze deelvraag 2, nl: *Beyond Memorability: Visualization Recognition and Recall* (Borkin et al., 2016) en *Measuring effectiveness of graph visualizations: A cognitive load perspective* (Huang, Eades, & Hong, 2009). Beide papers gaan in op *understandability* van visualisaties en hoe de cognitieve processen ervan worden gemeten waarbij ik bij de paper van Borkin et al. (2016) voor het eerst kennis maakte met de term *eye-tracking*. In deze paper las ik voor het eerst dat deze technologie wordt gebruikt om cognitieve processen in relatie tot visuele perceptie te meten. Deze term is daarom ook meegenomen in deelvraag 3.

Deelvraag 3: Hoe kan de relatie gelegd worden tussen intuïtie en de visuele perceptie van conceptuele modelleertalen?

Voor deze deelvraag zijn de volgende papers relevant:

Educating Intuition: A Challenge for the 21st Century (Hogarth, 2003) gevonden bij deelvraag 1 en

Beyond Memorability: Visualization Recognition and Recall (Borkin et al., 2016) gevonden bij deelvraag 2.

Measuring effectiveness of graph visualizations: A cognitive load perspective (Huang et al., 2009) gevonden bij deelvraag 2.

Uit de paper van Bork et. al (2019) *Intuitive Understanding of Domain-Specific Modeling Languages: Proposition and Application of an Evaluation Technique*, wordt Moody's Semantic Transparency genoemd als één van de design principes die een belangrijke rol speelt in de intuïteit van, een in dit geval, modelleer notatie. Door in deze paper backward snowballing te gebruiken kwam ik dan ook uit op de volgende relevante paper voor deze deelvraag: *The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering* (Moody, 2009).

Daarnaast zijn de volgende zoekwoorden gebruikt in Google Scholar: *modeling language AND eye-tracking*

Zoals aan het einde van de beschrijving van deelvraag 2 is aangegeven, wordt in de paper van Borkin et al. (2016) *eye-tracking* technologie gebruikt voor het meten van cognitieve processen aangaande de visuele perceptie tijdens het uitvoeren van taken. Daarom is in bovenstaande query *eye-tracking* opgenomen als zoekwoord.

De volgende relevante paper kwam hiermee naar boven: *Using-Eye-Tracking-to-Expose-Cognitive-Processes-in-Understanding-Conceptual-Models* van Bera, (Bera, Soffer, & Parsons, 2019) . In de bibliotheek van de OU vond ik geen relevante artikelen voor conceptuele modellen i.c.m. *eye-*

tracking. Door dezelfde query in Google Scholar in te vullen kwam ik op pagina drie het zesde item deze paper tegen. De paper geeft een uitgebreide beschrijving van het uitgevoerde experiment en wordt verder meegenomen in het onderzoek. Hiermee had ik een goed basis document gevonden om nog een paper erbij te zoeken. Door het toepassen van backward snowballing in deze paper kwam ik op de paper van Petrusel & Mendling (2013) met als titel: *Eye tracking the Factors of Process Model Comprehension Tasks*. Echter bleek na het bestuderen van het experiment dat de linkjes naar de procesmodellen die waren gebruikt niet meer werkten. Er ontbrak nogal wat informatie waardoor het niet mogelijk was een goed beeld te krijgen van het experiment. Deze paper viel af. Na weer backward snowballing toe te passen op de paper van Bera, kwam ik op de paper van Mendling & Recker (2008) met als titel: *Towards Systematic Usage of Labels and Icons in Business Process Models*. De paper bespreekt een systematische aanpak in het gebruik van pictogrammen en labels van activiteiten in een procesmodel en hoe deze kunnen bijdragen tot verbetering in het begrijpen van modellen. Uiteindelijk ben ik in het totaal uitgekomen op een zestal relevante papers.

2.2. Uitvoering

In deze paragraaf wordt de gevonden literatuur besproken. Vanwege de lengte van de analyses zijn deze te vinden in [bijlage 4](#).

2.3. Resultaten en conclusies

Conceptuele modelleertalen zijn in staat om complexiteit te verminderen en alleen de meest relevante zaken te tonen met als doel hierover met alle betrokkenen over te communiceren. Om te communiceren over een model speelt notatie een belangrijke rol, deze moet nl. de ontwerper van het model ondersteunen, maar ook degene die het model moet interpreteren (Bork et. al, 2019).

Deelvraag 1: Wat is intuïtie en waar bestaat deze uit?

Bork et al. (2019) spreekt over een intuïtieve notatie die moet zorgen voor het verminderen van de cognitieve belasting. Hogarth (2003) beschrijft twee denksystemen, nl. *deliberate* en het *tacit* systeem. Volgens Hogarth (2003) is met name dit laatste systeem verantwoordelijk van wat wij noemen "intuïtie". Wij mensen gebruiken onze zintuigen en bouwen door de jaren heen ervaringen op. Het *tacit* systeem zorgt ervoor dat al deze ervaringen en prikkels, samengevat als stimuli, die we binnenkrijgen tijdens ons leven onbewust worden opgeslagen in het langetermijngeheugen. De opslagcapaciteit van dit geheugen is onbeperkt en zorgt ervoor dat d.m.v. verbindingen wij als mensen kunnen begrijpen. Er is echter één nadeel aan dit geheugen en dat is dat het vrij traag is. Daarom bestaat er ook een kortetermijngeheugen of beter gezegd het werkgeheugen. Dit is sneller, maar heeft maar een beperkte opslagcapaciteit. Dit geheugen wordt gebruikt voor zaken die we maar kort hoeven te onthouden. Wij mensen zijn in staat om informatie tijdelijk uit het langetermijngeheugen te halen en deze te plaatsen in het kortetermijngeheugen wanneer we met die informatie direct aan de slag moeten. Het *deliberate* systeem is verantwoordelijk voor de bewuste processen, bijv. het lezen van een boek, een analyse maken, etc.. Deze processen vergen veel van onze resources. Het *tacit* en het *deliberate* systeem werken nauw samen, ook wanneer we

bepaalde beslissingen moeten nemen. Hogarth (2003) noemt hierin het voorbeeld met het besturen van een auto. Je bent bewust bezig met de auto een bepaalde richting uit te sturen, maar je hoeft niet na te denken bij het gas geven en het schakelen.

Het *tacit* denksysteem is dus verantwoordelijk van wat wij noemen 'intuïtie' wat zorgt voor onze onderbewuste denkprocessen dat informatie, afkomstig van onze zintuigen, opslaat in ons langetermijngeheugen. Deze informatie kunnen wij mensen aanspreken wanneer wij dat willen en zo het *tacit* systeem laten samenwerken met onze bewuste denk processen.

Deelvraag 2: Hoe kan intuïtie gemeten en gekwantificeerd worden?

Volgens Hogarth (2003) is de kwaliteit van onze 'intuïtie' gebaseerd op interacties/ervaringen die we hebben met de wereld om ons heen. Omdat denkprocessen, dus ook van het tacit denksysteem, zich vertaalt in acties, kunnen we stellen dat we in staat zijn de kwaliteit van intuïtie te operationaliseren door deze acties in een gecontroleerde omgeving te meten. De zgn. *comprehension tasks*, *problem solving tasks* of *cloze tests* worden vaak ingezet om m.b.v. deze acties de denkprocessen te meten waaronder ook het *tacit* systeem.

Het *tacit* systeem slaat ervaringen die we tijdens ons leven meemaken op in het langetermijngeheugen. Dit geheugen is belangrijk bij het begrijpen of herkennen van dingen. Borkin et al. (2016) geeft aan dat wanneer een observator een visualisatie niet begrijpt of herkent, te complex of onduidelijk is, hij/zij in het langetermijngeheugen blijft zoeken naar herkenbare punten. Wat dan opvalt is dat bij het uitvoeren van bijv. een *problem-solving task* de responstijd oploopt en dat dit invloed heeft op de score van het aantal correcte antwoorden. In de gevonden literatuur wordt naast het beantwoorden van vragen, *eye-tracking* ingezet om verschillende oogbewegingsparameters te registreren. In de uitslagen van eye-tracking metingen valt op dat wanneer een observator geen herkenbare aansluitpunten vindt in het langetermijngeheugen om een visualisatie te herkennen of te begrijpen, er op dat moment ook veel oogfixaties (één van de oogbewegingsparameters) worden geregistreerd. Borkin et al. (2016) en Bera et al. (2019) geven in hun experimenten een richting aan hoe het meten van begrijpelijkheid gekwantificeerd kan worden waarbij een techniek als *eye-tracking* vele deuren opent om cognitieve processen te meten.

Deelvraag 3: Hoe kan de relatie gelegd worden tussen intuïtie en de visuele perceptie van conceptuele modelleertalen?

Een model moet aan bepaalde eisen voldoen wil het correct geïnterpreteerd kunnen worden. Moody (2009) beschrijft in zijn "Physics" of Notations negen principes voor het ontwerpen van effectieve cognitieve visuele notatie. Mendling & Recker (2008) gaan in op het gebruik van tekstlabels in combinatie met pictogrammen en hoe deze samen kunnen leiden tot het bevorderen van begrijpelijkheid in een conceptueel model. De literatuur beschrijft dat indien een model geen duidelijke structuur heeft of informatie bevat die op meerdere manieren geïnterpreteerd kan worden dit de begrijpelijkheid van een model ondermijnt. Huang et al. (2009) toont in zijn experiment aan dat er vermoeidheid optreedt wanneer een model niet op een duidelijke manier zijn informatie toont waardoor de performance omlaag gaat van degene die het model moet interpreteren.

Wanneer een conceptueel model lastig is te interpreteren, gaat de cognitieve belasting omhoog en loopt de responstijd op. In zo'n situatie ontstaan er veel oogfixaties omdat de observator aan het

zoeken is naar herkenbare punten in de visualisatie die hij hoopt te vinden in het langetermijngeheugen (Borkin et al., 2016). Wanneer dit lang duurt zal er volgens Huang et al. (2009) een vorm van vermoeidheid optreden waardoor de performance omlaag gaat.

Visualisaties die herkenbaar en begrijpbaar zijn zullen een minder hoge cognitieve belasting tonen dan wanneer deze dat niet zijn. Borkin et al. (2016) gebruikt in hun experiment *eye-tracking* software om de data van de oogfixaties te verzamelen. Voor het verdere onderzoek zou *eye-tracking* ingezet kunnen worden om de cognitieve processen te meten die een rol spelen bij het kunnen begrijpen van een model.

2.4. Doel van het vervolgonderzoek

Aan de hand van de literatuuronderzoek wordt dit onderzoek vervolgd met het opzetten van een experimentele methode voor de vervolgevaluatie van de intuïtiviteit van de PGA modelleertaal. Nadien wordt een experiment uitgevoerd met een 10 tot twintigtal respondenten die zullen bestaan uit eerste en tweedejaars studenten Bedrijfskundige Informatica (BI). De verzamelde gegevens uit het experiment worden geanalyseerd waarbij de oude en nieuwe PGA notatie vergeleken wordt op intuïtiviteit. Ma.w. men wil in dit onderzoek een uitspraak kunnen doen of en in hoeverre de nieuw voorgestelde notatie intuïtiever is dan de oude notatie.

3. Methodologie

3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Het doel van dit onderzoek is om de intuïtiviteit van verschillende versies van de PGA notatie te evalueren.

Uit het experiment van Roelens & Bork (2020) zijn een aantal verbeterde PGA notaties voorgesteld die deels door menselijke creativiteit en inspiratie tot stand zijn gekomen. Het is echter niet vastgesteld of deze nieuwe notatie daadwerkelijk intuïtiever is dan de oude. In dit hoofdstuk zal antwoord gegeven worden op de volgende deelvraag:

Deelvraag 4: Hoe kunnen de gevonden resultaten uit het theoretisch kader vertaald worden in een concreet uitvoerbare methode?

Om dit te kunnen doen moet de gekozen onderzoeksmethode een deductieve aanpak ondersteunen. Volgens Saunders, Lewis, and Thornhill (2016) is een deductieve aanpak geschikt wanneer een theoretisch model in een gecontroleerde omgeving moet worden getest. In dit onderzoek wordt de theorie afkomstig uit de literatuurstudie uitgedrukt in hypothesen die de on- en afhankelijke variabelen beschrijven en wat de te verwachte resultaten zullen zijn. Om de hypothesen te toetsen is een experiment het meest geschikt vanwege het feit dat we (1) het experiment gaan doen met twee groepen (een *between subjects* opstelling) waarbij iedere groep een notatie versie krijgt, (2) om een statistisch causaal verband te vinden door na te gaan of een hypothese een verandering in één onafhankelijke variabele een verandering teweeg brengt in een andere afhankelijke variabele en (3) Het maakt het mogelijk om randomisatie toe te passen, m.a.w.

in dit onderzoek willen we de respondenten op willekeur plaatsen in één van de twee groepen (Saunders et al., 2016).

3.2. Technisch ontwerp: uitwerking van de methode

In de literatuur wordt een techniek als *eye-tracking* gebruikt binnen een experimentele opstelling om cognitieve processen op kwantitatieve wijze vast te leggen. “*Eye-tracking* stelt ons in staat te onderzoeken hoe cognitieve verwerking van informatie plaatsvindt en geeft inzicht in de inhoud van de te verwerken informatie” (Bera et al., 2019, p.3). Naast het gebruik van *eye-tracking* moeten respondenten ook zgn. *problem-solving tasks* en *comprehension questions* beantwoorden. I.c.m. *eye-tracking* data kan dit aangeven in hoeverre een model begrepen wordt door de lezer. De waarden die door *eye-tracking* kunnen worden afgeleid zijn: oogfixaties en daarvan de duur en het aantal keren de fixatie voorkomt en tot slot de saccades waarmee informatie kan worden verzameld over de zgn. *Areas Of Interest (AOIs)*. Naast de *eye-tracking* data wordt de responstijd opgenomen van de respondent wanneer deze een *comprehension of problem solving question* beantwoordt plus het aantal correcte antwoorden. Visualisaties waarbij de respondent geen herkenningspunten kan ontdekken zal de responstijd langer duren dan wanneer de visualisatie herkenbaar is. Als gevolg hiervan zal de score ook hoger zijn (Borkin et al., 2016). Tot slot wordt aan de respondenten gevraagd hoe zij het werken met de PGA-modelleertaal hebben ervaren. Hierbij vullen ze a.d.h.v. een vijftal meerkeuze vragen hun antwoorden in.

Dit leidt tot de volgende hypothesen:

H_a-fixatie duur: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal de algehele fixatie duur van de nieuwe PGA notatie lager zijn dan die van de oude PGA notatie

H₀-fixatie duur: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal de algehele fixatie duur van de nieuwe PGA notatie niet lager zijn dan die van de oude PGA notatie

Een hoge algehele fixatie duur wijst op het zoeken naar herkenbare punten die de respondent niet vindt. Dit zal voorkomen wanneer de visuele weergave niet duidelijk is en daardoor moeilijk te begrijpen.

H_a-fixatie aantal: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal het algehele fixatie aantal van de nieuwe PGA notatie lager zijn dan die van de oude PGA notatie

H₀-fixatie aantal: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal het algehele fixatie aantal van de nieuwe PGA notatie niet lager zijn dan die van de oude PGA notatie

Een hoge algehele fixatie aantal wijst op het zoeken naar herkenbare punten die de respondent niet vindt. Dit zal voorkomen wanneer de visuele weergave niet duidelijk is en daardoor moeilijk te begrijpen.

H_a-AOI fixatie duur: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal de AOI fixatie duur hoger zijn in de nieuwe PGA notatie dan bij de oude notatie.

H₀-AOI fixatie duur: Bij het uitvoeren van *comprehension* of *problem solving tasks*, zal de AOI fixatie duur niet hoger zijn in de nieuwe PGA notatie dan bij de oude notatie.

Wanneer een visualisatie wordt begrepen, worden de relevante gebieden ook goed gevonden en zullen die betreffende gebieden ook een hogere AOI fixatie duur hebben.

H_a-AOI run count: Bij het uitvoeren van *comprehension* of *problem solving tasks*, zal de AOI run count in de nieuwe PGA notatie hoger zijn in de relevante gebieden dan bij de oude notatie

H₀-AOI run count: Bij het uitvoeren van *comprehension* of *problem solving tasks*, zal de AOI run count in de nieuwe PGA notatie niet hoger zijn in de relevante gebieden dan bij de oude notatie

Wanneer een visualisatie wordt begrepen zal evenals bij de AOI fixatie duur de relevante gebieden goed gevonden worden en zullen die gebieden dan ook een hoge AOI run count hebben.

H_a-Comprehension questions effectiveness : Het aantal correct beantwoorde *comprehension questions* van de nieuwe PGA notatie is hoger dan die van de oude PGA notatie

H₀-Comprehension questions effectiveness : Het aantal correct beantwoorde *comprehension questions* van de nieuwe PGA notatie is niet hoger dan die van de oude PGA notatie

Bij een goed begrijpbare visuele weergave van de PGA model zullen de vragen correct beantwoord kunnen worden.

H_a-Problem solving effectiveness : Het aantal correct uitgevoerde *problem solving tasks* van de nieuwe PGA notatie is hoger dan die van de oude PGA notatie

H₀-Problem solving effectiveness : Het aantal correct uitgevoerde *problem solving tasks* van de nieuwe PGA notatie is niet hoger dan die van de oude PGA notatie

Bij een goed begrijpbare visuele weergave van het PGA model zullen de taken correct uitgevoerd kunnen worden.

H_a-Interpretational efficiency comprehension questions: De responstijd van de *comprehension questions* is in de nieuwe PGA notatie korter dan die van de oude PGA notatie

H₀-Interpretational efficiency comprehension questions: De responstijd van de *comprehension questions* is in de nieuwe PGA notatie niet korter dan die van de oude PGA notatie

Bij een goed begrijpbare visuele weergave van het PGA model zal de respondent in staat zijn sneller de relevante gebieden te vinden en dus sneller tot een antwoord kunnen komen.

H_a-Interpretational efficiency problem solving tasks: De responstijd van de *problem solving tasks* is in de nieuwe PGA notatie korter dan die van de oude PGA notatie

H₀-Interpretational efficiency problem solving tasks: De responstijd van de *problem solving tasks* is in de nieuwe PGA notatie niet korter dan die van de oude PGA notatie

Bij een goed begrijpbare visuele weergave van het PGA model zal de respondent in staat zijn sneller de relevante gebieden te vinden en dus sneller de taak zal uitvoeren.

H_a-Perceived ease of use : Het waargenomen gebruiksgemak van de nieuwe PGA notatie is hoger dan die van de oude PGA notatie.

H₀-Perceived ease of use: Het waargenomen gebruiksgemak van de nieuwe PGA notatie is niet hoger dan die van de oude PGA notatie.

Na afloop van het experiment zal aan de hand van een vragenlijst de respondenten gevraagd worden naar het gebruiksgemak van de notatie. Het resultaat van deze hypothese zal de bovenstaande hypothesen ondersteunen.

Wanneer respondenten de originele notatie en daarna de nieuwe notatie voorgelegd zouden krijgen, hebben we te maken met een *within-subjects* opstelling. Het probleem wat hierbij speelt, is de leerfactor die optreedt door de herhaling van experimentele taken. Dit laatste is niet aan de orde wanneer de respondenten in twee groepen worden opgedeeld (i.e. *between-subjects* opstelling). Het experiment zal een *between subjects* opstelling hebben om het leereffect zoveel mogelijk te vermijden.

De respondenten bestaan uit eerste- en tweedejaars studenten Bedrijfskundige Informatica (BI) van de Avans Hogeschool in Den Bosch die willekeurig, op basis van leeftijd en geslacht, worden onderverdeeld in twee groepen (A en B). De respondenten hebben ervaring met modelleren, maar zijn niet bekend met de PGA modelleertaal en zijn niet bekend met het domein van *strategic fit*. Voorafgaand aan het experiment ontvangen de respondenten instructies over hoe om te gaan met de *eye-tracking* software en de *questionnaire* en er wordt kort ingegaan op het doel van het onderzoek (zie [Bijlage 7](#)). Vanwege het Corona beleid wordt rekening gehouden dat het experiment naast een fysieke uitvoering ook online kan worden gehouden. Dit is mogelijk omdat de *eye-tracking* software en de *questionnaire* beide online services zijn die éénvoudig op afstand uitgevoerd kunnen worden waarbij de webcam gebruikt kan worden voor de *eye-tracking* service. Daarnaast worden de instructies en linkjes naar het experiment inclusief de unieke toegangscode per respondent via mail gecommuniceerd.

De respondent krijgt een mail (zie [Bijlage 7](#)) met daarin de linkjes naar het experiment wat bestaat uit vier secties, nl. (1) Demografische vragen, (2) *Comprehension Questions*, (3) *Problem Solving tasks* en (4) *Perceived ease of use*. Sectie 1 bestaat uit demografische vragen waar aan de respondent toestemming voor deelname wordt gevraagd. Kiest de respondent 'Nee' dan wordt het experiment afgebroken. Kiest hij/zij 'Ja' zal het experiment starten. Verder wordt in deze sectie gevraagd naar de ervaring die de respondent heeft op het gebied van modelleren. Secties 2 en 3 (i.e. *comprehension questions* en *problem solving tasks*) wordt met *eye-tracking* uitgevoerd. Sectie 4 bestaat uit de zgn. *Perceived ease of use* waar de respondent wordt gevraagd naar zijn ervaringen met de PGA-modelleertaal. De respondent krijgt de instructie dat hij/zij na het afronden van een sectie de volgende link mag aanklikken (zie [Bijlage 7](#) voor voorbeeld mail en aanwijzing in de *questionnaire*). Dit gaat zo door totdat alle vier de secties zijn afgerond pas dan is het experiment voltooid.

In de bijgeleverde instructie wordt de respondent ingelicht over hoe om te gaan met de *eye-tracking* software en er staat informatie over hoe het kalibreren werkt, hierbij wordt aangemerkt dat de respondent er voor moet zorgen dat hij/zij een werkende webcam/camera heeft om het experiment te kunnen doen. Het gehele experiment zal maximaal 30 minuten duren.

Experimentele taken

Fase 1: Case Study : *Comprehension questions*

Groep A krijgt een case study te zien in de oude notatie waarover een aantal *comprehension questions* worden gesteld.

Groep B krijgt dezelfde case study te zien, maar dan met de nieuwe notatie waarover een aantal *comprehension questions* worden gesteld.

Fase 2: Case Study : *Problem Solving tasks*

Groep A krijgen twee modellen te zien in de oude notatie waarbij de respondenten *problem solving tasks* moeten uitvoeren.

Groep B krijgen dezelfde modellen te zien in de nieuwe notatie waarbij de respondenten dezelfde *problem solving tasks* moeten uitvoeren.

Tot slot Fase 3: *Ease of use*

Aan de hand van een vragenlijst wordt aan het einde van het experiment gevraagd aan de respondenten hun ervaringen omtrent de modelleertaal te classificeren. De vragenlijst die hiervoor wordt gebruikt is dezelfde als die door Roelens & Bork (2020) wordt gebruikt.

3.3. Gegevensanalyse

De data die verzameld wordt, is onder te verdelen in continue en categorische data. Dit laatste betreft de antwoorden uit de vragenlijst die betrekking hebben op het gebruiksgemak van de notatie. De continue data bestaat uit (1) fixatie duur en aantal en saccadische data verzameld d.m.v. het toepassen van *eye-tracking*, (2) scores die worden behaald op de vragen of taken van de *problem solving tasks* en *comprehension questions* waarmee we de *effectiveness* kunnen meten en (3) responstijd verzameld van elk uitgevoerde *problem solving task* en *comprehension question* waarmee we de *efficiency* kunnen meten. De responstijd van elk *problem solving task* of *comprehension question* bestaat uit de duur vanaf het moment de vraag door de *questionnaire* wordt getoond tot het antwoord dat wordt gegeven door de respondent.

Omdat we te maken hebben met één populatie van studenten waaruit twee groepen zijn gehaald die afzonderlijk worden getest, wordt voor de continue variabelen de t-test gebruikt waarbij het gemiddelde van elke afhankelijke variabele in de twee ongepaarde groepen worden vergeleken. Op die manier kan worden aangetoond of de kans dat een verschil tussen de twee groepen alleen op toeval berust klein is. M.a.w. of deze statistisch significant is. De resultaten uit de analyse zullen worden besproken en getracht te worden verklaard. Om een t-test te gebruiken moeten de variabelen normaal verdeeld zijn. Om na te gaan of data een normale verdeling kent wordt de Shapiro-Wilk test gebruikt of een visualisatie als een Q-Q plot of histogram, mocht de Shapiro-Wilk

test niet overtuigend genoeg zijn. Categorische data kent van nature geen normale verdeling waardoor het gebruik van ongepaarde t-test niet zinvol is. Voor de categorische data wordt de ongepaarde Mann-Whitney-U test gebruikt. Deze test is ongevoelig voor niet-normaal verdeelde data.

Zoals al eerder vermeld dienen de afhankelijke continue variabelen een normale verdeling te hebben. Hiervoor zullen een aantal stappen genomen moeten worden, nl. de zgn. *outliers* zullen eerst geëlimineerd moeten worden omdat zij een vertekend gemiddelde kunnen geven aan de steekproeven. Daarnaast bestaat er de kans dat variabele als bijvoorbeeld tijd van de *eye-tracking* data, niet normaal verdeeld is. De variabelen zullen eerst getest moeten worden of deze normaal verdeeld zijn wat met behulp van de Shapiro-Wilk test of een visualisatie als Q-Q plots en histogram inzichtelijk kan worden gemaakt. De kans dat een normale verdeling niet vanzelfsprekend is heeft te maken met het feit dat niet alle oogbewegingen van de respondenten hetzelfde zijn. Het is daarom goed mogelijk dat een ideale distributie van oogmetingen niet altijd een normale verdeling laat zien. Respondenten met een lichte oogafwijking kunnen de algehele distributie verstoren. De *Between Subjects* zorgt er ook nog eens voor dat het aantal samples van iedere groep afzonderlijk niet groot genoeg is en daardoor de gevoeligheid op niet normaal verdeelde data groot kan zijn. Ook hier dient men rekening mee te houden in de analyses. Voor de transformatie naar normale verdeling kan bijv. een methode worden gebruikt als de *Log of Box-Cox transformation*, of door een alternatieve test als de Mann-Whitney U te gebruiken ingeval de data niet normaal verdeeld kan worden.

3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten

“Intuïtie is het resultaat van persoonlijke ervaringen” (Michael & Mayr, 2017, p. 2). Mensen zijn in staat onbewust te leren. De kwaliteit van dit leren bepaalt ook de kwaliteit van onze Intuïtie. Door dit proces zijn wij mensen in staat ‘dingen’ die wij waarnemen te begrijpen (Hogarth, 2003). Michael & Mayr (2017) halen nog een andere definitie aan voor intuïtie, nl. *understandability*. Er is dus een duidelijk verband tussen intuïtie en begrijpen.

Ethische aspecten

Het onderzoek zal volledig anoniem gebeuren. Wel wordt de respondent gevraagd naar de leeftijd en geslacht, dat enkel zal worden gebruikt om op een willekeurige wijze homogene groepen te creëren. Respondenten wordt niet gevraagd om antwoord te geven op hun fysieke gezondheid m.b.t. hun ogen. Dit gaat voorkomen worden door vooraf mede te delen dat alleen respondenten kunnen meedoen die geen bril hoeven te dragen tijdens het lezen van het beeldscherm i.v.m. het niet optimaal functioneren van de *eye-tracking* software. Via het interne systeem binnen de onderwijsinstelling zal een oproep worden geplaatst waar respondenten zicht kunnen aanmelden en waarop deze informatie te vinden is. Op die manier hoeven er geen medische gegevens opgeslagen worden.

Construct validiteit

In de papers lezen we dat visuele perceptie een belangrijke rol speelt in het begrijpen van dingen om ons heen en dat zij voor een groot deel verantwoordelijk is ons langetermijngeheugen te vullen om dingen in de wereld te begrijpen. Zo spelen ze ook een belangrijke rol in het begrijpen van een

modelleertaal. In de bestudeerde papers is veelal gebruik gemaakt van *eye-tracking* data omdat dit ons de mogelijkheid geeft via de ogen de bewegingen en afgelegde patronen vast te leggen als kwantitatieve data. Het geeft ons inzicht welke cognitieve processen er afspelen bij het interpreteren of begrijpen van een model. *Eye-tracking* data resulteert in twee soorten parameters, nl. fixatie en saccadische parameters waarmee de mate van het begrijpen van een model kan worden geïnterpreteerd. En zoals al eerder aangegeven is daarmee ook een verband gelegd naar het begrip intuïtie.

Betrouwbaarheid

Bij een *between subjects* wordt geen rekening gehouden met onderlinge verschillen binnen een groep, maar deze opzet zorgt er weer wel voor dat eventuele *outliers* zichtbaar worden. In de opzet van dit onderzoek, daar waar mogelijk, is rekening gehouden dat deze zoveel mogelijk in lijn is met het onderzoek van Roelens & Bork (2020). Hiermee bedoelen we de *between subjects* opzet, de nagenoeg identieke fase indeling en de *comprehension questions*, daarnaast zal de case study model (*heatmap*) hetzelfde zijn voor de originele notatie in één groep en zal in de andere groep hetzelfde model getoond worden maar dan met de nieuwe notatie. Dit is gedaan om zoveel mogelijk onvoorspelbare situaties binnen het experiment uit te sluiten en hierdoor zo min mogelijke bias te creëren bij de evaluatie.

In dit onderzoek wordt naast de *eye-tracking* data gebruik gemaakt van een combinatie van objectieve en subjectieve variabelen. Subjectieve variabelen zijn de resultaten afkomstig van de vragenlijst die gaan over het gebruikersgemak, deze worden in de analyse afgezet tegen de objectieve variabelen, de resultaten van de *comprehension questions* en *problem solving tasks* en hun responstijden. De triangulatie zorgt voor ondersteuning van de betrouwbaarheid waarbij het combineren van verschillende bronnen de resultaten op meerdere manieren kunnen worden belicht waarbij steeds hetzelfde verhaal als resultaat naar voren komt.

Interne validiteit

Eye-tracking data alleen is niet voldoende wanneer we niet weten waarom bepaalde oogbewegingen plaatsvinden. Fixatie aantal en duur bijvoorbeeld kunnen op meerdere manieren geïnterpreteerd worden. Het kan zijn dat de lezer interessante gebieden vindt, maar het kan ook duiden op verwarring. Variaties in *eye-tracking* data tonen aan dat ze afhankelijk zijn van bepaalde opdrachten (Bera et al., 2019). Daarom wordt de *eye-tracking* data gekoppeld aan *comprehension questions* en *problem solving tasks*. Bij de *comprehension questions* en *problem solving tasks* zijn de responstijden en het aantal correcte antwoorden of uitgevoerde taken van belang. De respondenten zijn verdeeld in twee groepen. Iedere groep heeft dezelfde of praktisch dezelfde leeftijdssamenstelling. De respondenten hebben dezelfde kennis niveau. In dit onderzoek worden de resultaten van het experiment onderworpen aan statistische testen. Wanneer blijkt dat de nieuwe notatie inderdaad intuïtiever is dan de oude, zal dit naar verwachting ook te terug te zien zijn in de verschillende resultaten tussen de twee groepen.

Verstorende factoren op interne validiteit

Verstorende factoren voor de interne validiteit zijn respondenten met een oogafwijking als bijv. een amblyopie (lui-oog) of bril, waarbij dit laatste niet zeker is of dit invloed heeft op de kwaliteit van *eye-tracking*. Om deze verstorende factoren uit te sluiten wordt gevraagd zich alleen maar aan te melden wanneer men geen bril draagt voor dichtbij of een oogafwijking heeft. Een andere verstorende factor is dat de respondent zijn hoofd zo stil mogelijk moet houden. Dit is bijna niet te voorkomen. In een aantal papers worden kin ondersteuning gebruikt, maar dit is logistiek niet te

regelen omdat vanwege de COVID-19 maatregelen de respondenten vanuit huis deelnemen aan het onderzoek. Wel kunnen er tips gegeven worden om het hoofd te ondersteunen, denk aan een stapel boeken waarop je je hoofd kunt laten rusten. Helaas is dit niet te controleren, dus dit kan een serieuze verstoring zijn. Vanwege de ongemakkelijke positie kan de duur van het experiment bepalend zijn voor de kwaliteit van het eindresultaat.

Vermoeidheid kan er voor zorgen dat respondenten willekeurige antwoorden geven. Voor een deel is dit op te vangen door rekening te houden met de duur van het experiment, het aantal vragen die worden gesteld, de opdracht omschrijvingen beknopt houden en tot slot het lettertype niet te klein maken. Een andere versturende factor kan zijn dat sommige respondenten dyslectisch zijn waardoor zij meer moeite hebben met lange stukken tekst waardoor zij overal meer tijd voor nodig hebben. In de vragenlijst die respondenten moeten invullen alvorens zij aan de opdrachten beginnen kan de vraag gesteld worden of en zo ja in hoeverre ze dyslexie hebben op een Likert schaal van 1 t/m 5.

De bestudeerde papers geven aan dat wanneer responstijd oploopt het niet altijd wil zeggen dat een respondent het lastig vindt het antwoorde te vinden in het model bij een *comprehension question*, het kan ook zijn dat het antwoord dat wordt gegeven door de respondent juist uitgebreid is omdat juist het model een hoge begrijpelijkheid heeft. In het onderzoek zal hier mee rekening gehouden moeten worden. Een manier om dit te voorkomen is om respondenten meerkeuze vragen te geven of door de vragen meer op te delen waardoor de antwoorden korter zijn en toch gedetailleerd genoeg.

Externe validiteit

Het experiment zal uit eerste- en tweedejaars studenten BI bestaan die geen ervaring hebben met de PGA notatie en het *strategic fit* domein, maar wel, zij het beperkt, kennis van conceptuele modelleertalen. Alhoewel de studenten een proxy zijn voor de werkelijke PGA eindgebruikers en dit experiment qua respondenten aansluit bij het onderzoek van Roelens & Bork (2020), zal deze keuze toch een beperkende invloed hebben op de generaliseerbaarheid van het experiment. Er zal meer onderzoek moeten plaatsvinden wil het experiment generaliseerbaar zijn voor de werkelijke PGA eindgebruikers binnen een bedrijfscontext.

4. Resultaten

4.1. Uitvoering

Respondenten waren eerste- en tweedejaars studenten Bedrijfskundige informatica aan de Avans hogeschool in Den Bosch. De test bestond uit vier onderdelen: (1) Demografische vragen, (2) *Comprehension Questions*, (3) *Problem Solving tasks* en (4) *Perceived ease of use*. Secties 2 en 3 (i.e. comprehension and problem solving tasks) werden met eye-tracking uitgevoerd. De test is tweemaal afgenomen vanwege de lage opkomst. De eerste keer was op 26 april en zijn er 34 studenten aangeschreven, de tweede keer was op 10 mei en zijn er 13 studenten benaderd. Vanwege het Corona beleid werd de eerste test online afgenomen. Aan de hand van een studentenlijst is een random sampling toegepast op basis van leeftijd die de studenten in twee groepen verdeelde, nl. één groep voor de oude notatie en één voor de nieuwe. De tweede poging kon fysiek worden gedaan. De studenten werd gevraagd of zij aan een onderzoek wilde meewerken waarbij ze, wanneer ze de test helemaal afrondten, een Subway bon konden ontvangen. In het totaal hebben er 21 studenten meegedaan aan het onderzoek.

Zoals al eerder aangegeven bestond het onderzoek uit 4 losse secties, dit had te maken met de online *eye-tracking* software. Secties 2 en 3 waren omwille van de kosten afgesteld tot maximaal 20 minuten. Dit tijdslot is vastgesteld door een proefpersoon met een soortgelijk profiel als dat van de respondenten de secties 2 en 3 te laten maken. De indeling van secties 2 en 3 voor de *eye-tracking* is als volgt: Sectie 2 voor iedere groep (oude notatie en nieuwe notatie) en datzelfde voor sectie 3 op die manier was het dan mogelijk om de secties 1 en 4 los van de *eye-tracking* te doen. Door de opnames en frames te bestuderen zijn er koppelingen kunnen leggen tussen de *eye-tracking* data en de antwoorden van de respondenten uit het enquête systeem voor de secties 2 en 3, resp. *comprehension questions* en *problem solving tasks*.

Het aantal respondenten wisselt erg sterk tussen de secties, daarom is hieronder een overzicht geplaatst met toelichting.

Tabel 4.1.1 Verdeling respondenten over de secties

Sectie 1	Sectie 2	Sectie 3	Sectie 4
21 Geen specifieke notatie verdeling	6 nieuwe notatie	8 oude notatie	18 Geen specifieke notatie verdeling
	6 oude notatie	7 nieuwe notatie	

In sectie 1 hebben alle aangemelde respondenten deelgenomen. Daarna zijn 3 respondenten gestopt na het afronden van sectie 1. Aan sectie 2, 3 en 4 hebben er in het totaal 18 respondenten meegedaan echter, wanneer de respondenten de *eye-tracking* data niet op de juiste manier afsloot, werd de data niet opgeslagen door de *eye-tracking* software. Het gevolg was dat die data niet gebruikt kon worden. Dit gebeurde een aantal malen en verklaart de wisselende bezetting tussen de secties.

4.2. Demografische gegevens

Er namen 21 studenten deel aan het onderzoek, waarvan 57% eerstejaars studenten. 42% gaf aan bekend of ervaring te hebben *Entity-Relationship modeling* waarbij er drie als opmerking gaven dat de ervaring kwam van de opdrachten van school. Vanaf het eerste jaar krijgen de studenten modelleren technieken als BPMN, EPC, UML activity diagrams, daarom werd de vraag gesteld of ze ervaring/bekend zijn met deze technieken. Deze vraag werd voor 90.5% met 'ja' beantwoord. Van alle respondenten gaf 66.5% een aanvulling bij de opmerkingen, zo gaf 38% aan bekend te zijn met *activity diagrams* en 28.5% bekend te zijn met BPMN. *Enterprise Architecture (EA)* is een keuze vak in het tweede jaar. Op de vraag of respondenten ervaring hebben of bekend zijn met *EA* gaf 71.4% aan hier niet mee bekend te zijn of ervaring mee te hebben. Eén respondent gaf aan ervaring te hebben met *Enhanced Entity- Relationship* modelleertechniek (*EER*) en *Data Flow Diagrams (DFD's)*. In tabel 4.2.1 staat een overzicht over de gemiddelde duur die werd besteed per sectie.

Tabel 4.2.1 Gemiddelde tijdbesteding per sectie

Sectie naam	Notatie versie	Gemiddelde duur in min. (M)	Respondenten (N)
demografische gegevens	Oud/Nieuw	1.23	21
<i>comprehension questions</i>	Oud	6.19	6

<i>comprehension questions</i>	Nieuw	7.0	6
<i>problem solving tasks</i>	Oud	4.5	8
<i>problem solving tasks</i>	Nieuw	5.58	7
<i>perceived ease of use</i>	Oud/Nieuw	0.8	18

Er is door de respondenten gemiddeld 14.61 minuten besteed aan de test met de nieuwe notatie en 12.72 minuten aan de oude notatie.

4.3. Algemene resultaten

Tabel 4.3.1 Overzicht resultaten hypothese toetsen

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
Algehele fixatie aantal	<i>comprehension questions</i>	.90	.20	6 oude notatie	5.17 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.83 nieuwe notatie		
Algehele fixatie duur	<i>comprehension questions</i>	.925	.150	6 oude notatie	5.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	8.00 nieuwe notatie		
Algehele fixatie aantal	<i>problem solving tasks</i>	.947	.105	8 oude notatie	6.25 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	10.00 nieuwe notatie		
Algehele fixatie duur	<i>problem solving tasks</i>	.975	.049	8 oude notatie	5.88 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	10.43 nieuwe notatie		
AOI fixatie duur	<i>comprehension questions</i>	.261	.522	6 oude notatie	5.83 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.17 nieuwe notatie		
AOI fixatie run count	<i>comprehension questions</i>	.0325	.065	6 oude notatie	4.58 oude notatie	Verwerpen	Aanvaarden
				6 nieuwe notatie	8.42 nieuwe notatie		
AOI fixatie duur	<i>problem solving tasks</i>	.0245	.049	8 oude notatie	5.88 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	10.43 nieuwe notatie		
AOI fixatie run count	<i>problem solving tasks</i>	.003	.006	8 oude notatie	5.06 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	11.36 nieuwe notatie		
		.601	.799	6 oude notatie	6.75 oude notatie	Aanvaarden	Verwerpen

<i>Comprehension questions effectiveness</i>	<i>comprehension questions</i>			6 nieuwe notatie	6.25 nieuwe notatie		
<i>Problem solving effectiveness</i>	<i>problem solving tasks</i>	.093	.186	8 oude notatie	6.63 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.57 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	<i>comprehension questions</i>	.789	.423	6 oude notatie	5.76 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.33 nieuwe notatie		
<i>Interpretational efficiency problem solving tasks</i>	<i>problem solving tasks</i>	.877	.247	8 oude notatie	6.75 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.43 nieuwe notatie		
<i>perceived ease of use</i>	<i>perceived ease of use</i>	.016	.032	8 oude notatie	7.36 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	12.86 nieuwe notatie		

Tabel 4.3.1 laat ons een overzicht zien van de resultaten afkomstig van de hypothese toetsen. Alle toetsen zijn gedaan met de Mann-Whitney U test. In de meeste gevallen gaf de normaal test van Shapiro-Wilk aan dat de data een niet-normale verdeling had. In sommige gevallen was ($p > .05$), maar wanneer de histogrammen en Q-Q-plots werden geraadpleegd was de data niet-normaal verdeeld, daarnaast is de steekproefgrootte vrij klein waarvan bekend is dat niet-parametrische testen het beter doen met kleine steekproeven dan de parametrische testen. Op basis van histogrammen, Q-Q plots en de kleine steekproefgrootte werd besloten om de Mann-Whitney U test toe te passen. De histogrammen en Q-Q plots zijn opgenomen in [Bijlage 5](#). Verder wordt veelvuldig gebruikt gemaakt van de effectgrootte. Door de effectgrootte te berekenen van de uitkomst van een niet-parametrische test kan men zien hoe groot de proportie aan variatie is in een afhankelijke variabele dat verklaard kan worden door de onafhankelijke variabele. De formule en de tabel die hiervoor wordt gebruikt kan eveneens teruggevonden worden in [Bijlage 6](#).

In de volgende paragrafen worden de resultaten meer in detail toegelicht.

4.4. Hypotheses H-fixatie duur en H-fixatie aantal

H_a-fixatie aantal: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal de algehele fixatie aantal van de nieuwe PGA notatie lager zijn dan die van de oude PGA notatie

H₀-fixatie aantal: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal de algehele fixatie aantal van de nieuwe PGA notatie niet lager zijn dan die van de oude PGA notatie

H_a-fixatie duur: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal de algehele fixatie duur van de nieuwe PGA notatie lager zijn dan die van de oude PGA notatie

H₀-fixatie duur: Bij het uitvoeren van de *comprehension of problem solving tasks*, zal de algehele fixatie duur van de nieuwe PGA notatie niet lager zijn dan die van de oude PGA notatie

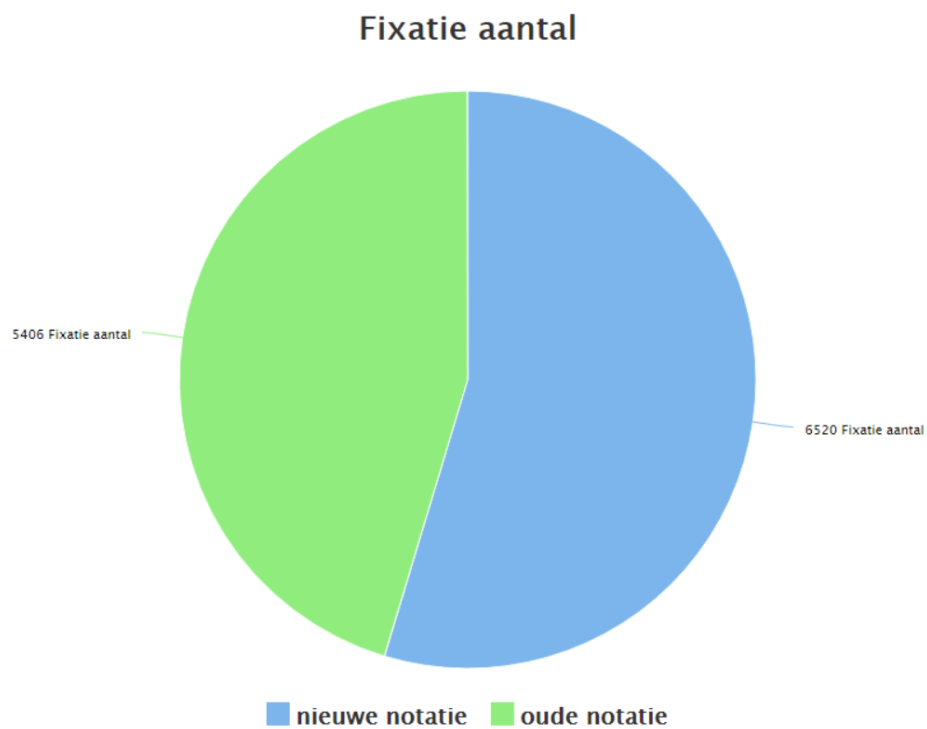
Om bovenstaande hypothesen te testen gebruiken we van de *eye-tracking* data het attribuut *fixation duration*. De variabele meet de duur van een fixatie in milliseconden (ms). Gedurende de hele *eye-tracking* test van een respondent wordt zijn fixatie duur gemeten. Dit wordt gedaan voor zowel de oude als de nieuwe notatie. Voor de data preparatie wordt verwezen naar [Bijlage 5](#).

Tabel 4.4.1 Overzicht resultaten hypothese toetsen algehele fixatie duur en aantal

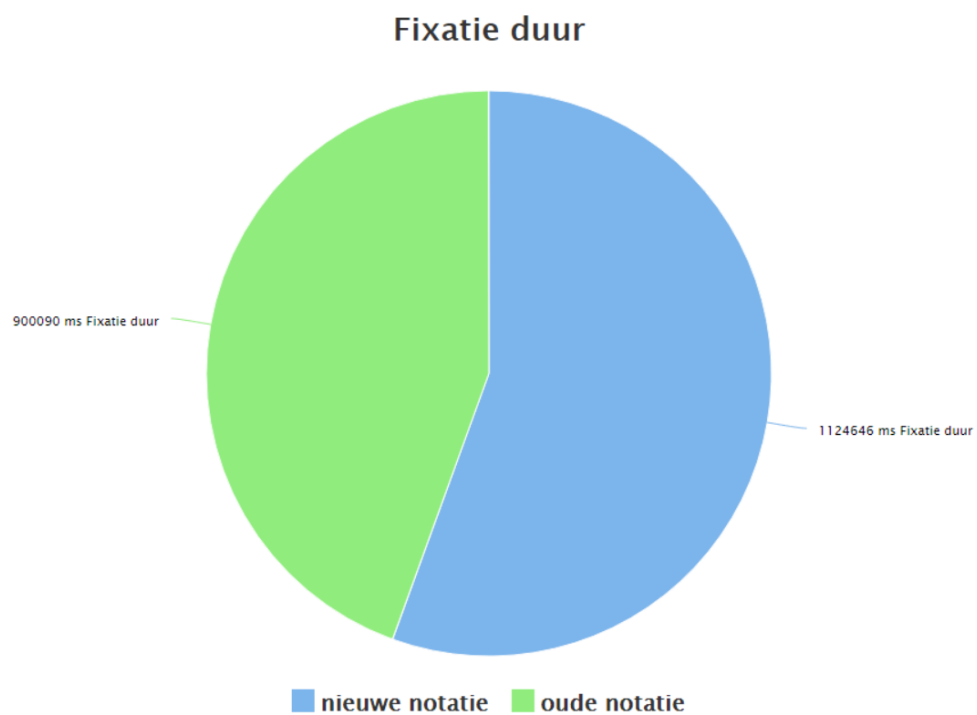
Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
Algehele fixatie aantal	<i>comprehension questions</i>	.90	.20	6 oude notatie	5.17 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.83 nieuwe notatie		
Algehele fixatie duur	<i>comprehension questions</i>	.925	.150	6 oude notatie	5.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	8.00 nieuwe notatie		
Algehele fixatie aantal	<i>problem solving tasks</i>	.947	.105	8 oude notatie	6.25 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	10.00 nieuwe notatie		
Algehele fixatie duur	<i>problem solving tasks</i>	.975	.049	8 oude notatie	5.88 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	10.43 nieuwe notatie		

Tabel 4.4.1 laat zien dat de *Mean Ranks (MR)* in zowel bij de *comprehension questions* als bij de *problem solving tasks* de nieuwe notatie hoger scoort terwijl we zouden verwachten dat de nieuwe notatie lager zou scoren dan de oude notatie. De *comprehension questions* laten een verschil zien tussen de *MRs* van $(7.83 - 5.17) = 2.66$ voor de algehele fixatie aantal en voor de algehele fixatie duur een verschil van $(8 - 5) = 3$. De verschillen in de *MRs* voor de *problem solving tasks* zijn als volgt, $(10.00 - 6.25) = 3.75$ voor de algehele fixatie aantal en $(10.43 - 5.88) = 4.55$ voor de algehele fixatie duur. De grootste verschillen tussen de oude en nieuwe notatie zien we in de algehele fixatie duur van de *problem solving tasks* terug.

De figuren 4.4.2a en b tonen de algehele fixatie aantallen en duur tussen oude en nieuwe notatie van de *comprehension questions*.



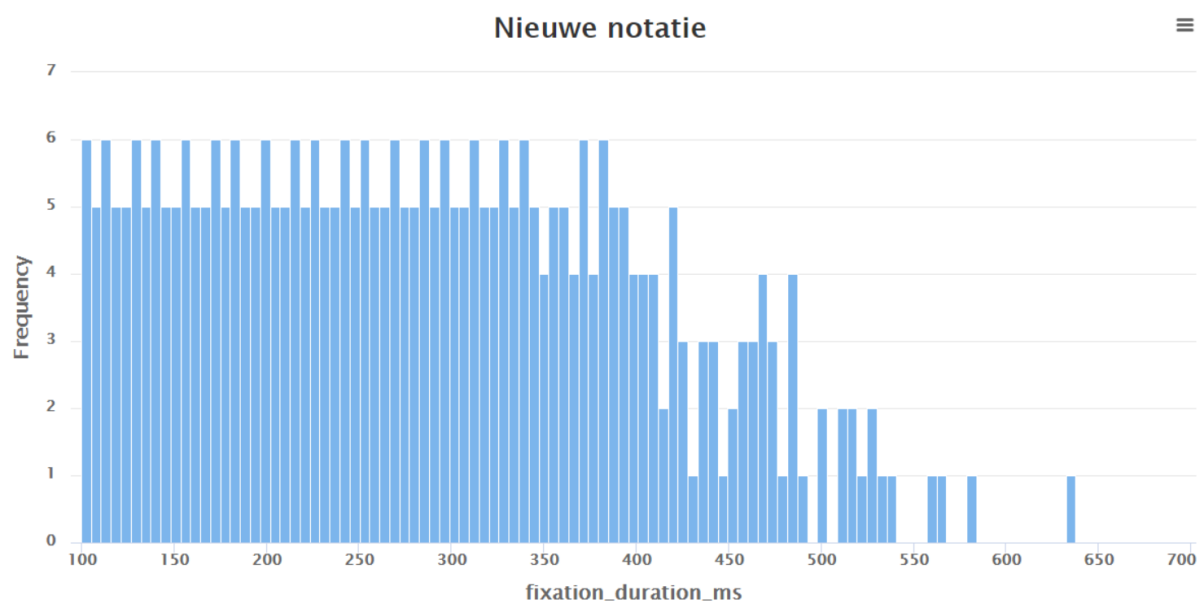
Figuur 4.4.2a: *comprehension questions* fixatie aantal



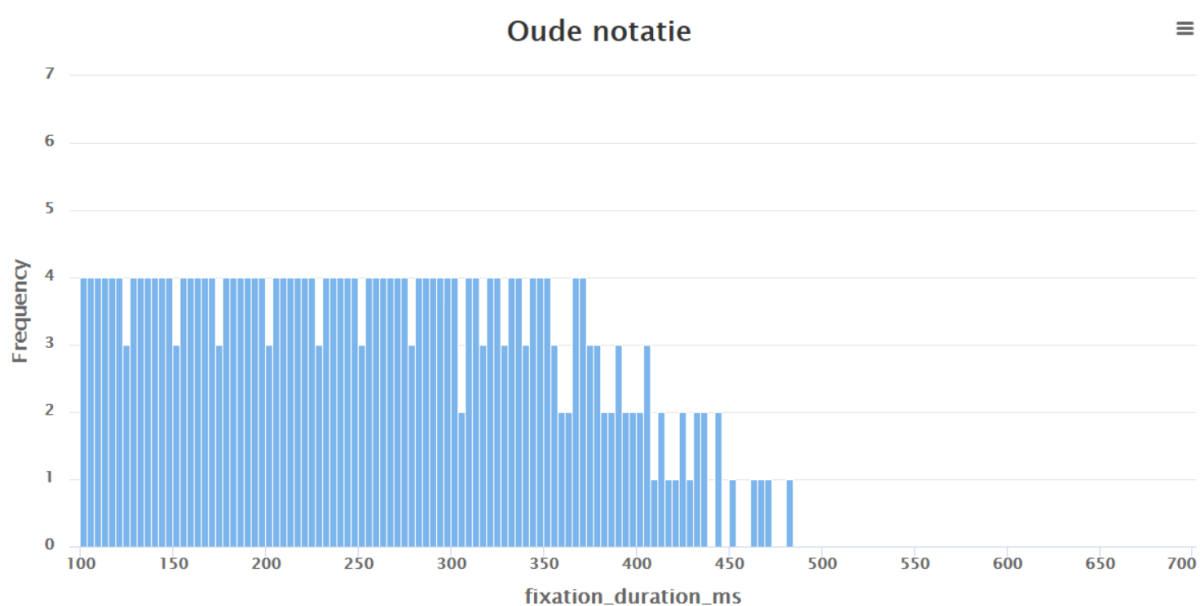
Figuur 4.4.2b: *comprehension questions* fixatie duur

In figuur 4.4.2a wordt het totaal aantal gemaakte fixaties in een pie-chart weergegeven. Deze zijn als volgt: de nieuwe notatie heeft 6520 fixaties met $N = 6$, N is het aantal respondenten, en de oude 5406 met $N = 6$. Dit is in de nieuwe notatie een toename van $(6520 - 5406) / 5406 = 0.2 = 20\%$

Figuur 4.4.2b geeft een overzicht van de totale duur van de fixaties van beide notatieversies. Ook hier heeft de nieuwe notatie een veel hogere fixatie duur dan de oude notatie, resp. 1124646 ms bij $N = 6$ en 900090 ms bij $N = 6$. Dit is een toename in de nieuwe notatie van ca. 24%. Dit zien we terug in de *MR* verschillen.

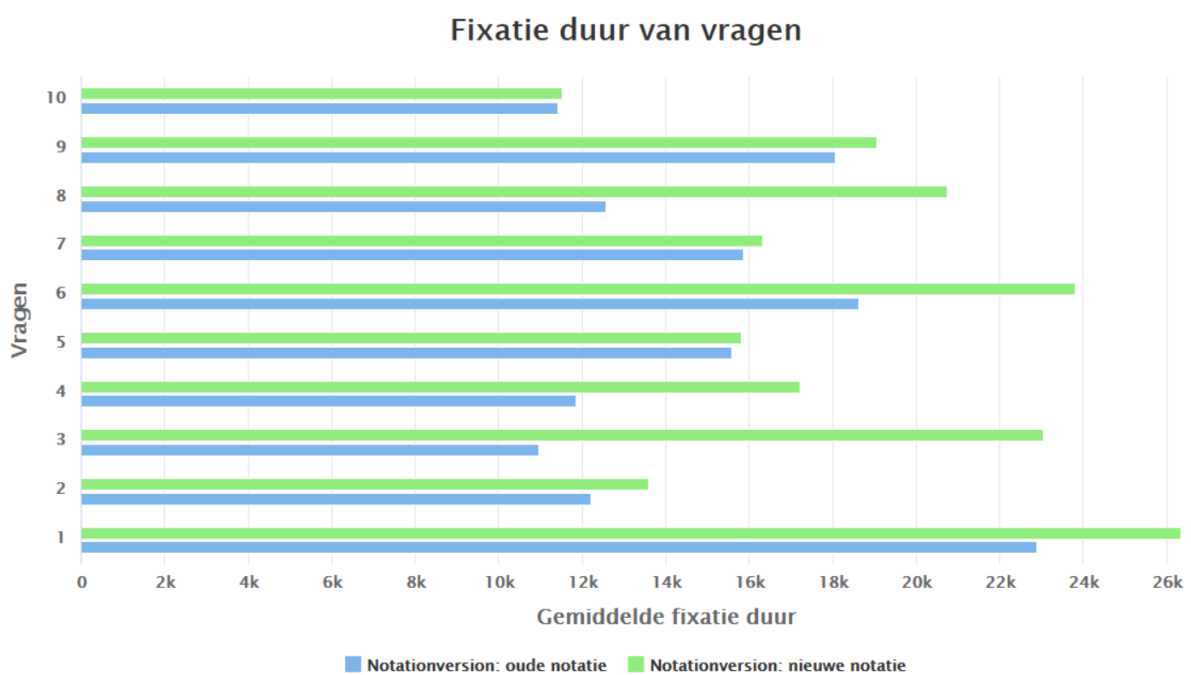


Figuur 4.4.3a *comprehension questions* frequentie van fixatie duur van nieuwe notatie

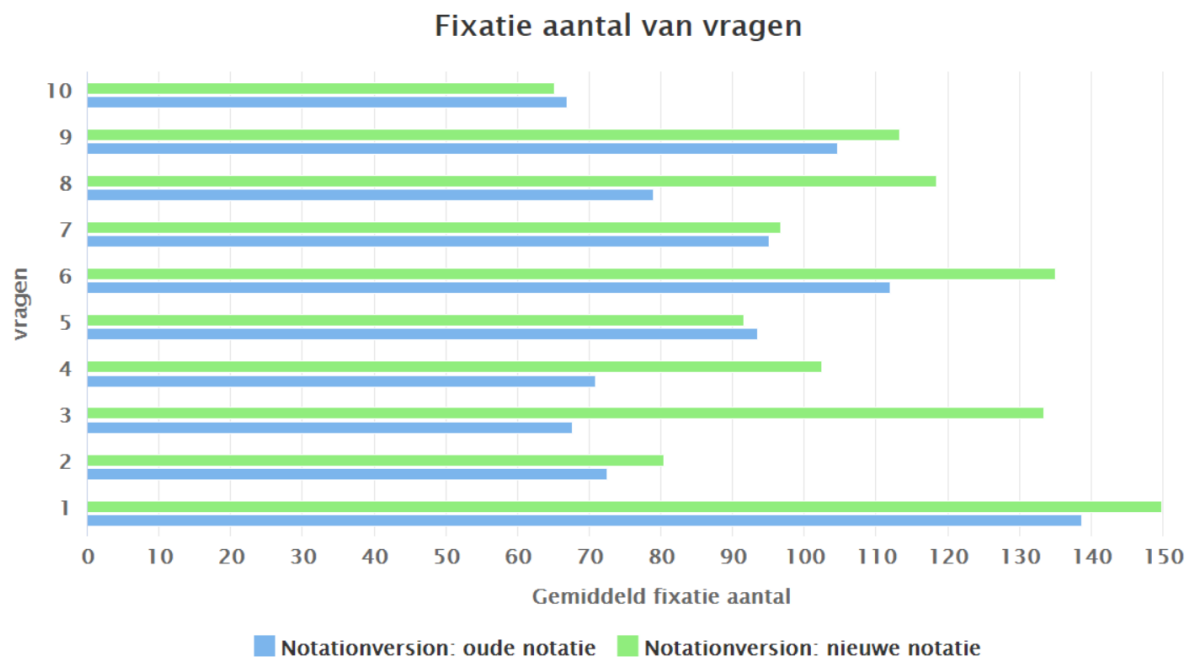


Figuur 4.4.3b *comprehension questions* frequentie van fixatie duur van oude notatie

In bovenstaande figuren 4.4.3a en b is te zien dat de fixatieduur in de oude notatie begint terug te lopen tussen de 350 ms en 480 ms waarbij er vanaf de 400 ms het merendeel bestaat uit éénmalige fixaties. Bij de nieuwe notatie nemen de aantal fixaties af van ongeveer 400 ms maar blijven dan nog voor een merendeel rond een aantal van 2 tot 3 fixaties hangen. Er wordt langer gekeken naar de modellen van de nieuwe notatie. De oogfixaties duren in de nieuwe notatie langer en de aantal oogfixaties liggen vanaf de ca. 400 ms hoger dan in de oude notatie.



Figuur 4.4.4a *comprehension questions* fixatie duur van vragen

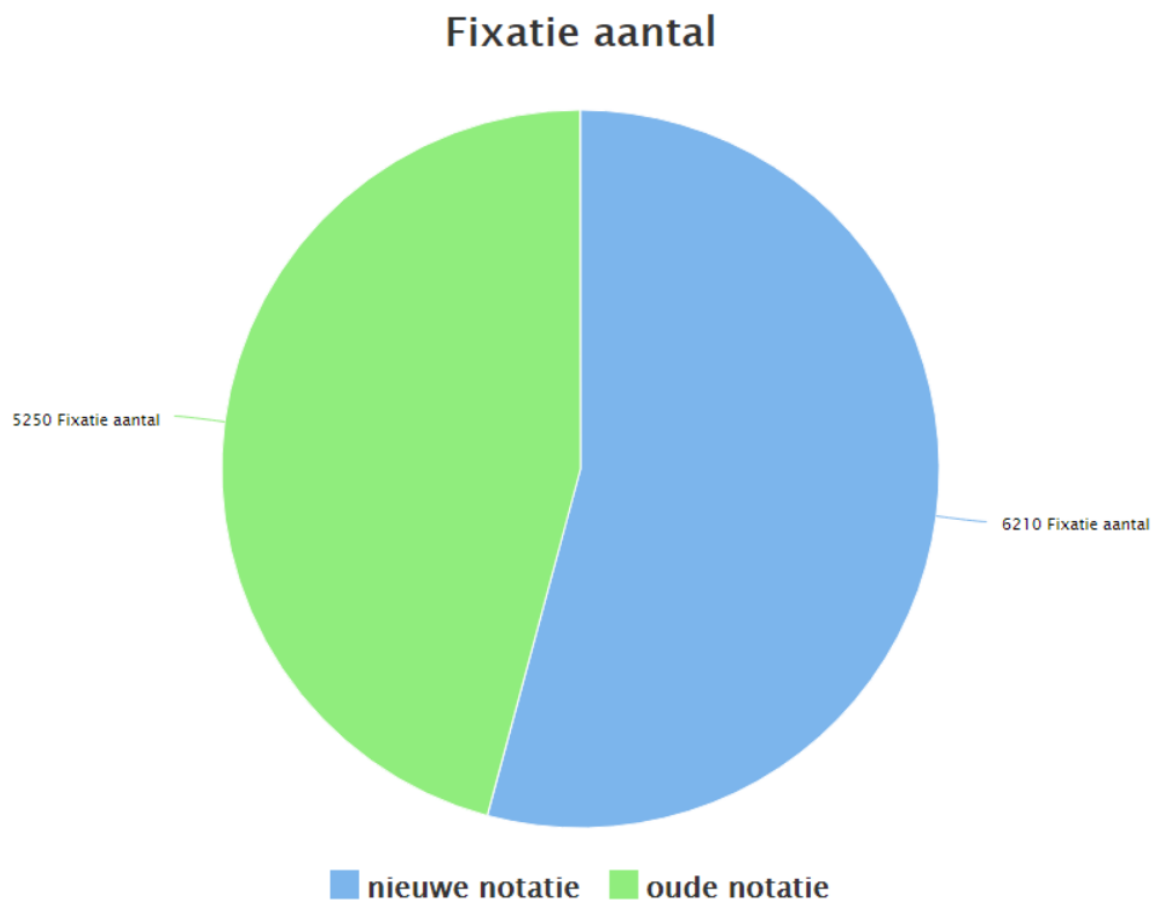


Figuur 4.4.4b *comprehension questions* fixatie aantal van vragen

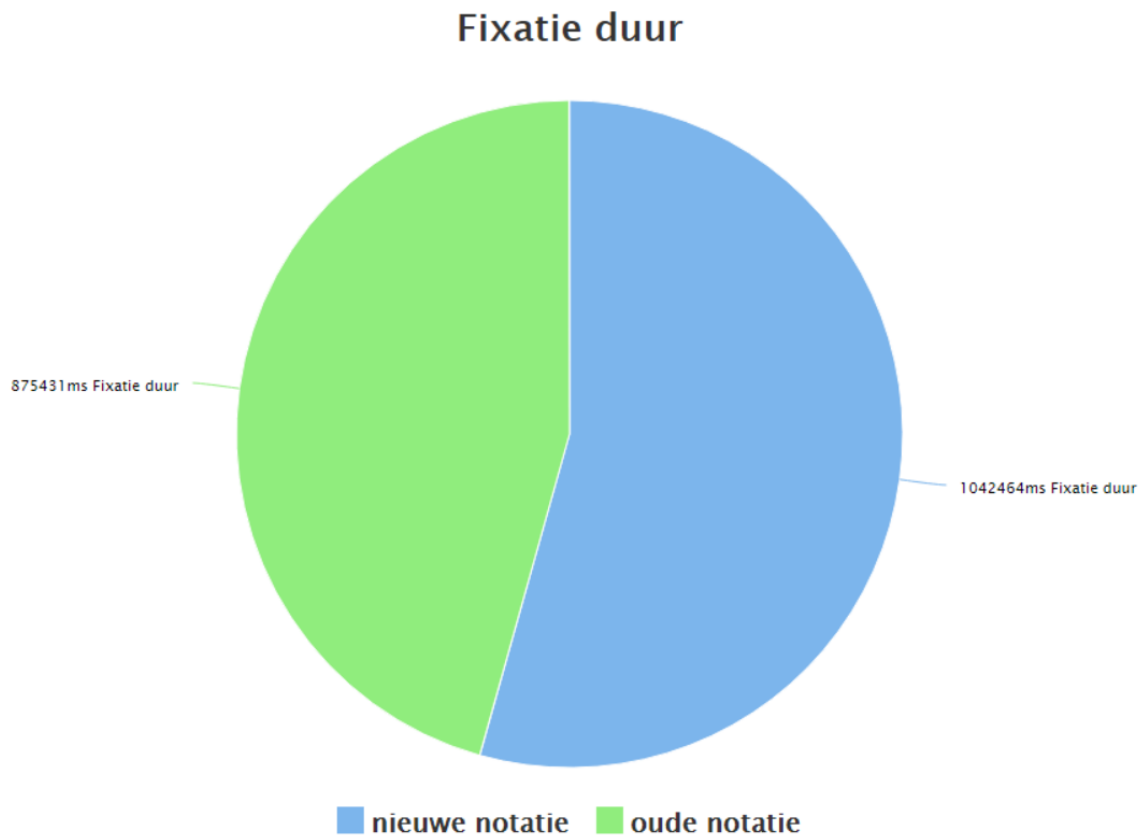
Tabel 4.4.4c *comprehension questions* scores

vragen versie	1	2	3	4	5	6	7	8	9	10
Oud	2	0	1	2	2	0	1	3	4	0
Nieuw	3	0	1	5	3	0	0	1	1	1

Figuren 4.4.4a t/m c tonen de gemiddelde fixatie aantallen en duur van de vragen en de totaal behaalde score per vraag. De vragen 1, 4, 5 en 10, dit gaat over de elementen: *customer goal*, *valueStream-importance*, *valueStream* en *competence*, scoren in de nieuwe notatie hoger dan de oude (zie tabel 4.4.4c), maar zij tonen allemaal een hogere fixatie duur in de nieuwe notatie t.o.v. de oude (zie figuur 4.4.4a). In figuur 4.4.4b hebben de vragen 5 en 10 (*valueStream* en *competence*) een lager fixatie aantal in de nieuwe notatie dan bij de oude en weer een hoog fixatie aantal bij de vragen 1 en 4. Men verwacht dat de nieuwe PGA notatie een lagere fixatie aantal en duur zal hebben dan de oude notatie, vanwege het feit dat de visualisatie beter begrepen wordt en men hierdoor minder hoeft te zoeken. Dit resultaat zien we niet terug, daarom zal de hypothese H_a verworpen en H_0 aanvaard worden voor de fixatie duur en aantal.



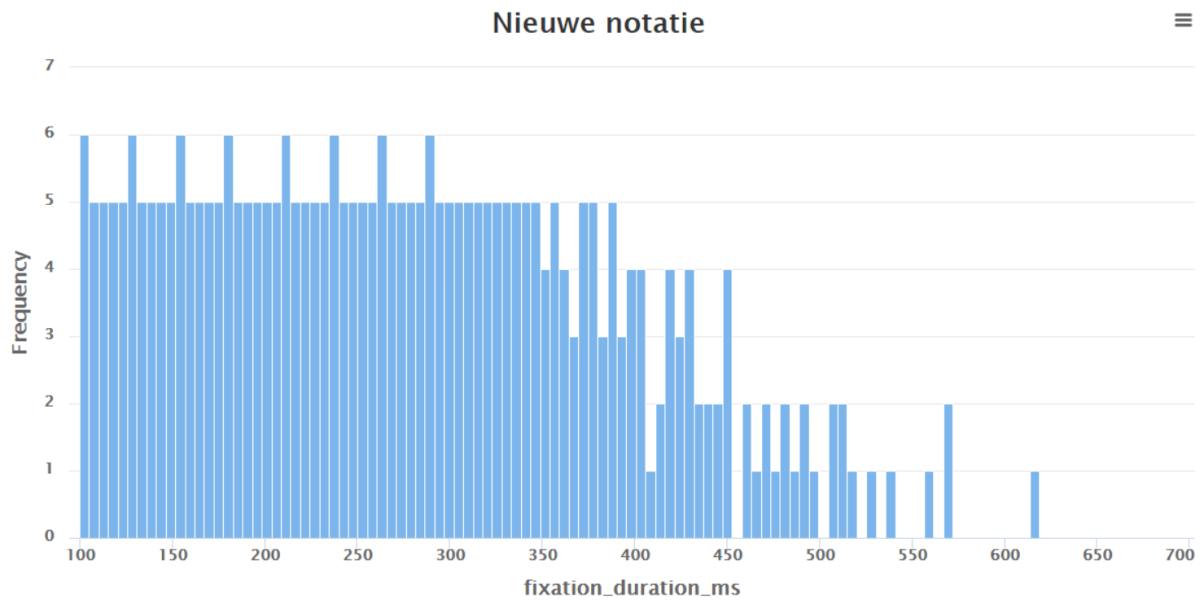
Figuur 4.4.5a : *problem solving tasks* fixatie aantal



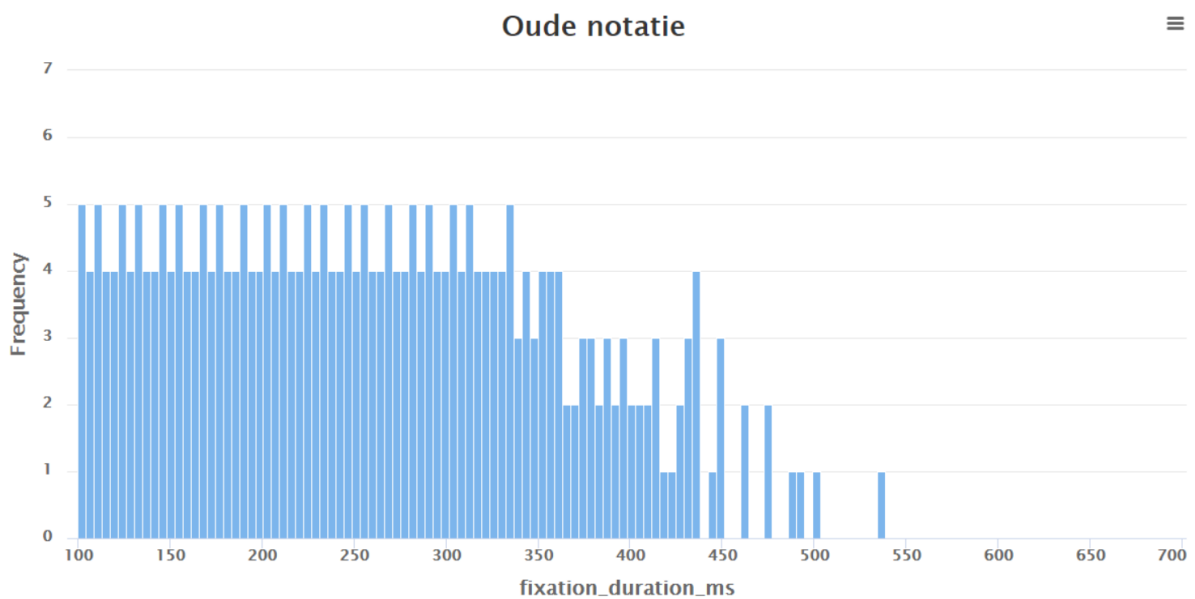
Figuur 4.4.5b : *problem solving tasks* fixatie duur

Figuren 4.4.5a en b laten het totaal aantal fixaties en duur zien tussen beide notaties bij de *problem-solving tasks*. De nieuwe notatie heeft een totaal aantal fixaties van 6210 met $N = 7$, N is het aantal respondenten en 5250 met $N = 8$ voor de oude notatie. Vanwege de ongelijke aantallen respondenten wordt de totale fixatie duur en aantal teruggebracht naar één respondent: 887.14 en 656.25 voor resp. de nieuwe en oude notatie. Het verschil hiertussen is 230.89 wat uitkomt op een toename van 35 %.

Totale duur van de fixaties in de nieuwe notatie is 1042464 met $N = 7$ en in oude notatie 875431 met $N = 8$. Deze cijfers tonen nagenoeg hetzelfde beeld. Visueel wijken de verhoudingen tussen de *pie-charts* in figuren 4.4.5a en 4.4.5b niet sterk af. De nieuwe notatie heeft, teruggebracht naar één respondent, een duur van 148923.42 ms, bij de oude is dat 109428.87 ms. Dit is een verschil van 39494.55 ms wat neer komt op een toename van 36 %.

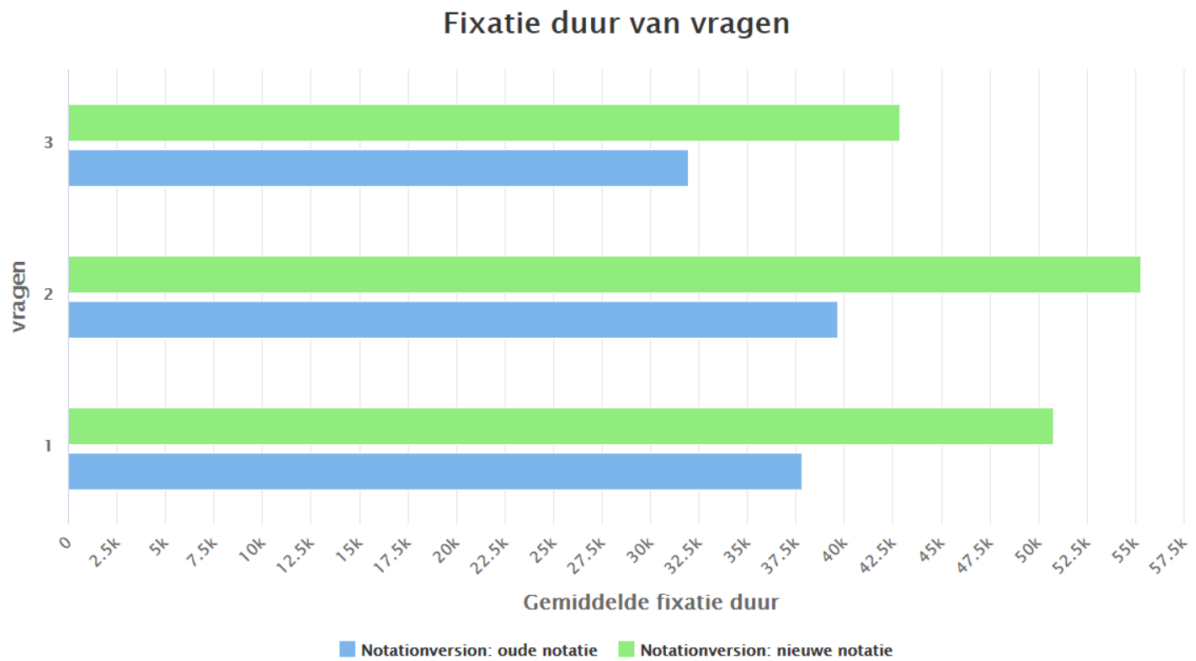


Figuur 4.4.6a *problem solving tasks* frequentie van fixatie duur van nieuwe notatie

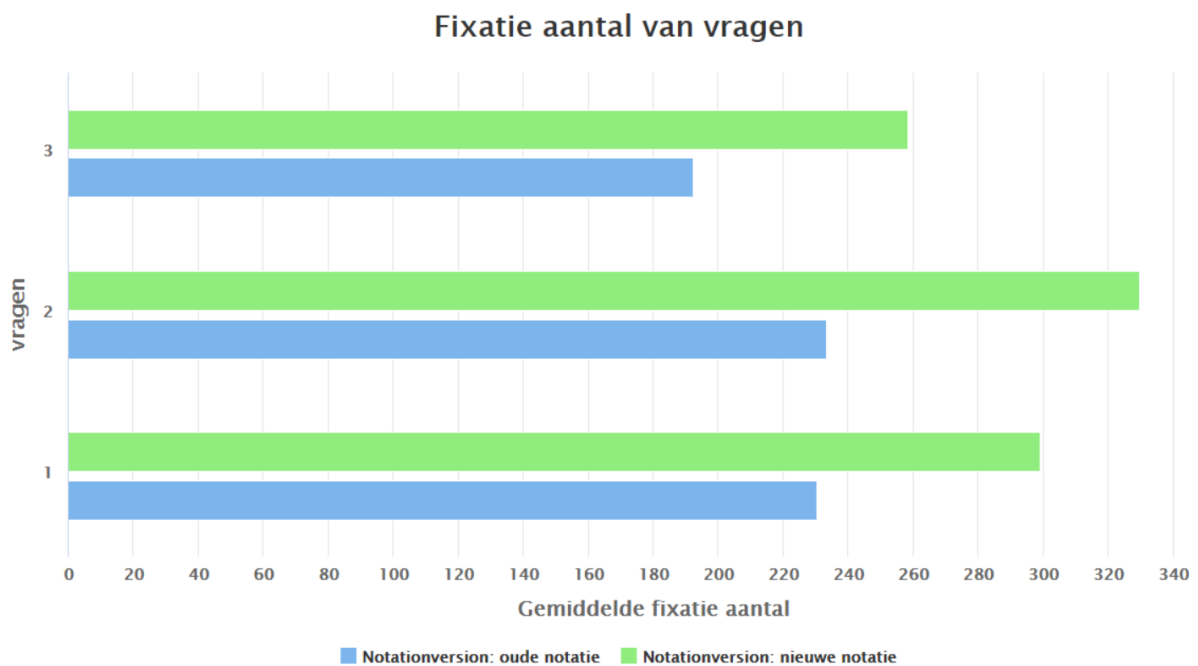


Figuur 4.4.6b *problem solving tasks* frequentie van fixatie duur van oude notatie

In bovenstaande figuren 4.4.6b is te zien dat de fixatieduur in de oude notatie begint terug te lopen tussen de 330 ms naar 480 ms. Anders dan bij de *comprehension questions* houden de *problem solving tasks* in beide notaties langer hogere fixatie aantallen vast bij hoge fixatie duur. De nieuwe notatie heeft een lange staart vanaf de ca. 400 ms en blijft ook relatief iets langer rond de 2 fixaties hangen dan de oude notatie.



Figuur 4.4.7a *problem solving tasks* fixatie duur van vragen



Figuur 4.4.7b *problem solving tasks* fixatie aantal van vragen

Tot slot tonen de figuren 4.4.7a en b aan dat de algehele fixatie duur en aantal in de nieuwe notatie hoger scoort dan de oude in de *problem solving tasks*. Daarmee geldt ook voor deze taak dat H_a wordt verworpen en H_0 wordt aanvaard.

4.5. Hypotheses H-AOI fixatie duur en H-AOI *run count*

H_a-AOI fixatie duur: Bij het uitvoeren van *comprehension* of *problem solving tasks*, zal de AOI fixatie duur hoger zijn in de nieuwe PGA notatie dan bij de originele notatie.

H₀-AOI fixatie duur: Bij het uitvoeren van *comprehension* of *problem solving tasks*, zal de AOI fixatie duur niet hoger zijn in de nieuwe PGA notatie dan bij de originele notatie.

H_a-AOI run count: Bij het uitvoeren van *comprehension* of *problem solving tasks*, zal de AOI run count in de nieuwe PGA notatie hoger zijn in de relevante gebieden dan bij de originele notatie

H₀-AOI run count: Bij het uitvoeren van *comprehension* of *problem solving tasks*, zal de AOI run count in de nieuwe PGA notatie niet hoger zijn in de relevante gebieden dan bij de originele notatie

Om bovenstaande hypothesen te toetsen gebruiken we van de *eye-tracking* data de attributen *aoi_fixation_average_duration_ms*, *aoi_fixation_total_count*. Voordat deze variabelen aan een significantie toets onderworpen kunnen worden, zijn er vooraf een aantal data preparatie stappen genomen. Deze zijn terug te vinden in [Bijlage 5](#).

Tabel 4.5.1 Overzicht resultaten hypothese toets *comprehension questions* AOI fixatie duur en *run count*

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
AOI fixatie duur	<i>comprehension questions</i>	.261	.522	6 oude notatie	5.83 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.17 nieuwe notatie		
AOI fixatie run count		.0325	.065	6 oude notatie	4.58 oude notatie	Verwerpen	Aanvaarden
				6 nieuwe notatie	8.42 nieuwe notatie		

Voor de *comprehension questions* scoort de AOI fixatie duur hoger in de nieuwe notatie dan in de oude, maar heeft een *p*-waarde van .261 (zie tabel 4.5.1) en is daarmee niet significant ($p > .05$). Hetgeen betekent dat H_a wordt verworpen en H₀ aanvaard, dus dat het waargenomen effect statistisch niet significant is bevonden ondanks dat we zien dat de nieuwe notatie hoger scoort dan de oude notatie. De AOI fixatie *run count* wat een *p*-waarde heeft van .032 een *MR* = 4.58 voor de oude en een *MR* = 8.42 voor de nieuwe notatie, wordt significant bevonden waardoor we H₀ kunnen verwerpen en H_a aanvaarden.

Tabel 4.5.2 *comprehension questions* per vraag AOI *run count*

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
AOI fixatie <i>run count</i>	Vraag 1 (<i>Customer goal</i>)	.030	.060	6 oude notatie	4.75 oude notatie	Verwerpen	Aanvaarden
				6 nieuwe notatie	8.25 nieuwe notatie		
AOI fixatie <i>run count</i>	Vraag 2 (<i>Valueproposition</i>)	.467	.934	6 oude notatie	6.42 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.58 nieuwe notatie		
AOI fixatie <i>run count</i>	Vraag 3 (<i>Internal goal</i>)	.351	.702	6 oude notatie	6.17 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.83 nieuwe notatie		
AOI fixatie <i>run count</i>	Vraag 4 (<i>valueStream-Importance</i>)	.052	.104	6 oude notatie	4.83 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	8.17 nieuwe notatie		
AOI fixatie <i>run count</i>	Vraag 6 (<i>valueStream-Importance-competence</i>)	.842	.317	6 oude notatie	7.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.00 nieuwe notatie		
AOI fixatie <i>run count</i>	Vraag 7 (<i>Goals-Valueproposition</i>)	0.266	.531	6 oude notatie	5.92 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.08 nieuwe notatie		
AOI fixatie <i>run count</i>	Vraag 8 (<i>valueStream-Valueproposition</i>)	0.200	.400	6 oude notatie	5.83 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.17 nieuwe notatie		

Tabel 4.5.2 toont de MRs per vraag van de AOI *run count*. De vragen 9 (*valueStreams-Importance-performance*) en 10 (*competence*) vonden in beide versie geen AOI gebieden, dit geldt ook voor vraag 5. Deze laatste ging over de *valueStream*, dit element was te klein voor de *eye-tracking* software om te detecteren. Er zat nl. een kleine afwijking in de software waardoor de fixaties soms naast de gebieden vielen. Wat in de opname wel duidelijk naar voren kwam was dat respondenten bij deze vraag de *mousepointer* gebruikte om de *valueStreams* te tellen.

De grootste verschillen in de nieuwe t.o.v. de oude notatie zien we bij de vragen 1 en 4 deze gaan over de elementen *Customer goal* en *valueStream-Importance*. Het element *Customer goal* is met de *p*-waarde van .030 en een *MR* = 4.74 in de oude en een *MR* = 8.25 in de nieuwe notatie significant bevonden dat op de voet wordt gevolgd door het element *valueStream-Importance* met een *p*-waarde van .052 een *MR* = 4.83 in de oude en een *MR* = 8.17 in de nieuwe notatie. De *p*-waarde ligt net boven het significantie niveau en wordt dus als niet significant aangemerkt. De derde variabele is die van vraag 8 die betrekking heeft op de elementen *valueStream-Value proposition*. De *p*-waarde is

.200 met een $MR = 5.83$ in de oude notatie en een $MR = 7.17$ in de nieuwe. De MR van de nieuwe notatie ligt op 7.17 waardoor deze is $> 50\%$ en dus het waargenomen effect neigt naar de alternatieve hypothese H_a . Kijken we naar de effectgrootte van deze drie variabelen dan zien we dat de vragen 1 en 4 resp. een effectgrootte hebben van ($r = .54$) 54% wat een groot effect is en ($r = .47$) 47% een middelmatig effectgrootte. Daarentegen is de effectgrootte van vraag 8 ($r = .24$) 24% wat een kleine effectgrootte is en wat aangeeft dat de afhankelijke variabele nauwelijks tot geen invloed heeft.

De algemene test geeft aan dat de nieuwe notatie hoger scoort dan de oude waarbij de kans dat deze op toeval berust erg klein is. Uit de resultaten blijkt dat de vragen 1 en 4 duidelijk beter scoren in de nieuwe notatie en dat daarvan vraag 1 een significant resultaat geeft. Beide vragen, betrekking hebbend op de elementen *Customer goal*, *valueStream-Importance*, hebben resp. een middelmatig tot groot effect op de afhankelijke notatieversie variabele. We kunnen stellen dat het significante resultaat uit de algemene test wordt bepaald door deze elementen. Conclusie H_0 wordt verworpen en H_a wordt aanvaard.

Tabel 4.5.3 Overzicht resultaten hypothese toets *problem solving tasks* AOI fixatie duur en *run count*

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H_0	H_a
AOI fixatie duur	<i>problem solving tasks</i>	.0245	.049	8 oude notatie	5.88 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	10.43 nieuwe notatie		
AOI fixatie run count		.003	.006	8 oude notatie	5.06 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	11.36 nieuwe notatie		

Voor de *problem solving tasks* scoort de AOI fixatie duur beter in de nieuwe notatie ($MR = 10.43$) met een p -waarde van .0245 (zie tabel 4.5.3) en geeft daarmee een significant resultaat weer. Dit betekent dat H_0 wordt verworpen en H_a aanvaard. Hetzelfde geldt voor de AOI fixatie *run count* met een p -waarde van .003 wat betekent dat H_0 wordt verworpen en H_a wordt aanvaard.

In de tabellen 4.5.4a en b zijn de resultaten per vraag vastgelegd en verdeeld in de AOI fixatie *run count* en duur. Het eerste wat opvalt is dat vraag 1 in beide hypothese toetsen voor de *problem solving tasks* het minst significant resultaat geeft ($p = .357$ en $p = .403$) t.o.v. de andere vragen ondanks dat de MR in de nieuwe notatie hoger is dan de oude. Wat ook opgemerkt dient te worden is dat vraag 1 het beste scoort in de oude notatie t.o.v. de overige vragen van de oude notatie. Bij vraag 1 werd aan de respondenten gevraagd de fout te ontdekken in het model. In het model moest de *value proposition* icoon vervangen worden door de *competence* icoon.

In tabel 4.5.4a (AOI fixatie duur) is te zien dat van de drie vragen vraag 3 (Onkosten efficiënter afhandelen) een significant resultaat ($p = .013$) oplevert met een $MR = 5.63$ in de oude notatie en $MR = 10.71$ in de nieuwe notatie wat een verschil oplevert van 5.08. Vraag 2 (Te hoge baktijd) geeft net geen significant resultaat, maar scoort ook hoger in de nieuwe notatie dan in de oude. Dit is terug te zien in de verschillen tussen de MR s, deze is nl. 3.75. De effectgrootte van de vragen 2 en 3

zijn resp. ($r = 0.42$) 42% wat een middelmatig effectgrootte is en ($r = 0.57$) 57% wat een groot effectgrootte is.

Tabel 4.5.4b laat de resultaten zien van de AOI *run count*. Anders dan bij de AOI fixatie duur vinden we bij AOI *run count* twee significante resultaten, het betreft hier vraag 2 (te hoge baktijd) ($p = .0075$) en vraag 3 (onkosten efficiënter afhandelen) ($p = .016$) met resp. $MR = 5.38$ in de oude en $MR = 11.00$ in de nieuwe notatie en $MR = 5.69$ in de oude notatie en $MR = 10.64$ in de nieuwe notatie. De effectgrootte van vraag 2 (te hoge baktijd) is ($r = .63$) 63%, wat aangeduid kan worden als een groot effect. Vraag 3 (onkosten efficiënter afhandelen) heeft een effectgrootte van ($r = .56$) 56%, wat eveneens aangeduid kan worden als een groot effect.

Tabel 4.5.4a *problem solving tasks* per element AOI fixatie duur

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
AOI fixatie duur	Vraag 1 (<i>value proposition</i> vervangen door <i>competence</i>)	.357	.714	8 oude notatie	7.63 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	8.43 nieuwe notatie		
AOI fixatie duur	Vraag 2 (te hoge baktijd)	.052	.105	8 oude notatie	6.25 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	10.00 nieuwe notatie		
AOI fixatie duur	Vraag 3 (Onkosten efficiënter afhandelen)	.013	.027	8 oude notatie	5.63 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	10.71 nieuwe notatie		

Tabel 4.5.4b *problem solving tasks* per element AOI fixatie *run count*

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
AOI fixatie <i>run count</i>	Vraag 1 (<i>value proposition</i> vervangen door <i>competence</i>)	.403	.807	8 oude notatie	7.75 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	8.29 nieuwe notatie		
AOI fixatie <i>run count</i>	Vraag 2 (te hoge baktijd)	.0075	.015	8 oude notatie	5.38 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	11.00 nieuwe notatie		

AOI fixatie <i>run count</i>	Vraag 3 (Onkosten efficiënter afhandelen)	.016	.032	8 oude notatie	5.69 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	10.64 nieuwe notatie		

De algemene test geeft aan dat de AOI fixatie duur voor de *problem solving tasks* in de nieuwe notatie hoger scoort dan de oude waarbij de kans dat deze op toeval berust erg klein is. Wanneer we de afzonderlijke vragen gaan testen op significantie, zien we dat er twee vragen hoog scoren en dat vraag 1 het minst scoort. Ondanks dat vraag 2 ook hoog scoort is de score niet significant te noemen, dat kan alleen worden toegeschreven aan vraag 3. De AOI fixatie *run count* heeft twee significante resultaten in de vragen 2 en 3 en ook hier scoort vraag 1 het laagst. We kunnen stellen dat H_a fixatie duur en H_a fixatie *run count* beide worden aanvaard en dat H_0 fixatie duur en H_0 fixatie *run count* beide worden verworpen.

Tot slot nog een algemene analyse waarbij we de totale duur van fixaties en het totaal aantal fixaties afzetten tegen de AOI fixatie duur en *run count*, zien we het volgende.

Tabel 4.5.5 totale fixatie duur en aantal afgezet tegen AOI fixatie duur en *run count*

Taak	Notatie versie	Totale fixatie duur in ms	Totale AOI fixatie duur in ms	Totaal aantal fixaties	Totaal AOI fixatie <i>run count</i>	AOI fixatie duur % t.o.v. totale duur	AOI <i>fixatie run count</i> % t.o.v. totaal <i>run count</i>
<i>comprehension questions</i>	Oud	900090	11071	5406	73	1.2	1.3
<i>comprehension questions</i>	Nieuw	1124646	17237	6520	177	1.5	2.7
<i>problem solving tasks</i>	oud	875431	31934	5250	252	3.6	4.8
<i>problem solving tasks</i>	Nieuw	1042464	48209	6302	440	4.6	6.9

De kolommen in tabel 4.5.5 'AOI fixatie duur % t.o.v. totale duur' en 'AOI fixatie *run count* % t.o.v. totaal *run count*' geven het aandeel AOI fixatie *run count* en AOI fixatie duur in percentages weer t.o.v. de algehele fixatie duur en aantal. Dus hoeveel procent van de algehele fixatie duur en aantal is weg te schrijven naar resp. AOI fixatie duur en *run count*.

In eerste instantie merken we op dat over het algemeen de AOI fixatie duur en *run count* hoger is bij de *problem solving tasks* dan bij de *comprehension questions*. Dit is te verklaren omdat de *comprehension questions* vaak direct vraagt naar elementen of verbindingen van en naar elementen. Het resultaat is dan ook dat de respondent direct naar het relevante AOI gebied kijkt behorende bij de vraag. Bij de *problem solving tasks* is dit los gelaten en vragen we bij deze taak een probleem te ontdekken waardoor de respondent meer op zoek moet gaan naar de relevante gebieden.

Wanneer de oude en nieuwe notaties van de *comprehension questions* worden vergeleken uit tabel 4.5.5, zien we een verschil tussen nieuwe notatie en oude notatie van $(1.5\% - 1.2\%) = 0.3\%$ voor de AOI fixatie duur. Voor de AOI fixatie aantal is dit $(2.7\% - 1.3\%) = 1.4\%$. Voor de *problem solving tasks* noteren we een verschil van $(4.6\% - 3.6\%) = 1\%$ en $(6.9\% - 4.8\%) = 2.1\%$ in resp. AOI fixatie duur en AOI fixatie aantal. Kijken we bijv. naar de nieuwe notatie van de *problem solving tasks*, dan zien we dat 4.6% van de totale fixatie duur die plaatsvindt weggeschreven kan worden naar de AOI fixatie duur en 6.9% is het aandeel AOI fixatie *run count* t.o.v. de totale fixatie aantallen. Vergelijken we deze percentages met de uitkomsten van de oude notatie van deze taak dan zien we dat de nieuwe notatie in verhouding hoger scoort t.o.v. de oude notatie.

4.6. Hypotheses H-Comprehension questions effectiveness en H-Problem solving effectiveness

H_a-Comprehension questions effectiveness : Het aantal correct beantwoorde *comprehension questions* van de nieuwe PGA notatie is hoger dan die van de originele PGA notatie

H₀-Comprehension questions effectiveness : Het aantal correct beantwoorde *comprehension questions* van de nieuwe PGA notatie is niet hoger dan die van de originele PGA notatie

Voor de data preparatie wordt verwezen naar [Bijlage 5](#).

Tabel 4.6.1 toont het resultaat van de algemene test voor de *comprehension questions effectiveness* en *problem solving effectiveness*

Tabel 4.6.1 Overzicht resultaten hypothese toets *comprehension questions effectiveness*

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
Comprehension questions effectiveness	comprehension questions	.601	.799	6 oude notatie	6.75 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.25 nieuwe notatie		

Tabel 4.6.1 toont aan dat het aantal correcte antwoorden in de *comprehension questions* in de nieuwe notatie niet leiden tot een statistische significant resultaat.

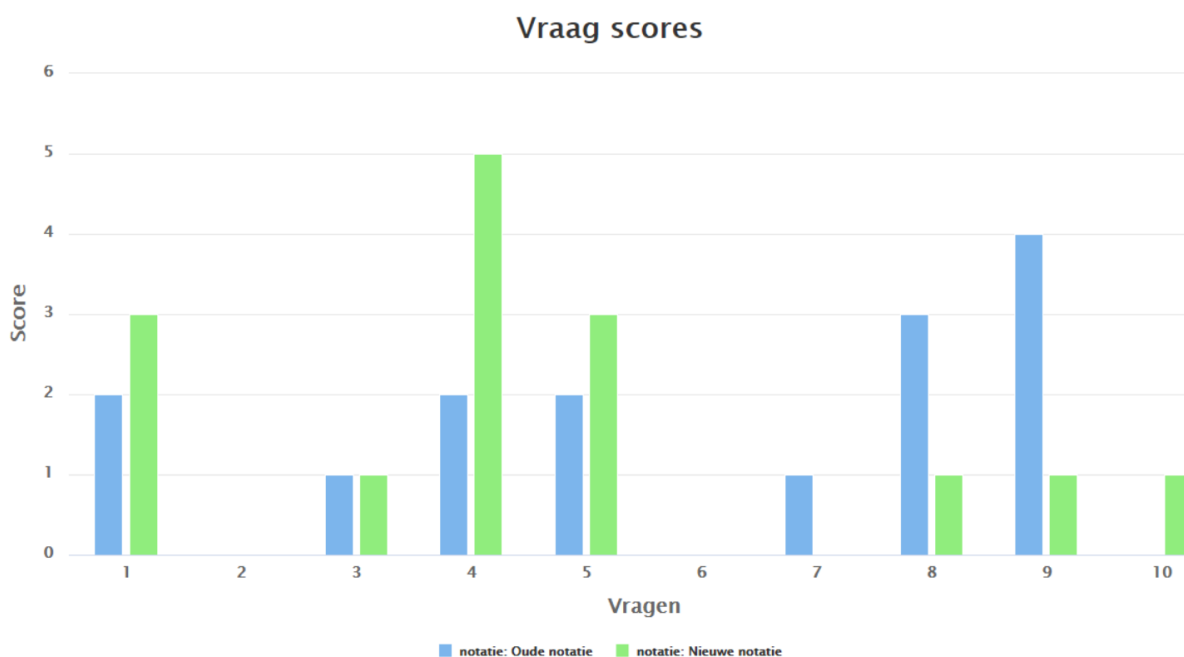
In het hierna volgende gedeelte wordt er dieper ingegaan op de vragen van de *comprehension questions*.

Tabel 4.6.2 *comprehension questions effectiveness* per vraag

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
Comprehension questions effectiveness	Vraag 1 (Customer goal)	.287	.575	6 oude notatie	6.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.00 nieuwe notatie		
Comprehension questions effectiveness	Vraag 2 (Valueproposition)	.500	1.000	6 oude notatie	6.50 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.50 nieuwe notatie		
Comprehension questions effectiveness	Vraag 3 (Internal goal)	.500	1.000	6 oude notatie	6.50 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.50 nieuwe notatie		
Comprehension questions effectiveness	Vraag 4 (valueStream-Importance)	.046	.093	6 oude notatie	5.00 oude notatie	Verwerpen	Aanvaarden
				6 nieuwe notatie	8.00 nieuwe notatie		
Comprehension questions effectiveness	Vraag 5 (valueStream)	.287	.575	6 oude notatie	6.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.00 nieuwe notatie		
Comprehension questions effectiveness	Vraag 6 (valueStream-Importance - competence)	.500	1.000	6 oude notatie	6.50 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.50 nieuwe notatie		
Comprehension questions effectiveness	Vraag 7 (Goals-Valueproposition)	.842	.317	6 oude notatie	7.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.00 nieuwe notatie		
Comprehension questions effectiveness	Vraag 8 (valueStream-Valueproposition)	.880	.241	6 oude notatie	7.50 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	5.50 nieuwe notatie		
Comprehension questions effectiveness	Vraag 9 (valueStreams-Importance - performance)	.954	.093	6 oude notatie	8.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	5.00 nieuwe notatie		

Comprehension questions effectiveness	Vraag 10 (competence)	.158	.317	6 oude notatie	6.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.00 nieuwe notatie		

Ondanks de resultaten uit de algemene test die aangeeft H_0 te aanvaarden en H_a te verwerpen, zien we in tabel 4.6.2 dat vraag 4 (gaat over het element *valueStream-Importance*) statistisch significant wordt bevonden.



Figuur 4.6.3 *comprehension questions* vragen afgezet tegen scores

We zien in figuur 4.6.3 vier vragen (1, 4, 5 en 10), waarbij de nieuwe notatie beter scoort dan de oude. De *MRs* zijn $> 50\%$ waardoor het waargenomen effect gaat in de richting van de alternatieve hypothese H_a . De vragen 1, 4, 5 en 10 gaan over resp. *Customer Goal*, *valueStream-Importance*, *valueStream* en *Competence*. Bij de vragen 2 en 6 (*value proposition* en *valueStream-Importance-competence*) wordt door de respondenten in beide notatie versies een score van 0 behaald waardoor ze niet zichtbaar zijn in de figuur. De nieuwe notatie scoort het laagst t.o.v. de oude in vraag 9. Bij deze vraag moeten de respondenten het kritiek pad in het model vinden. Daarbij moeten ze de *valueStreams*, *Importance* en *performance* elementen kunnen interpreteren. Deze vraag heeft een $MR = 5$ en een p -waarde van .954. Vraag 4, *valueStream-Importance* heeft een $MR = 8.00$ en een p -waarde van .046 waarmee dit element statistisch significant is. De effectgrootte van deze vraag is ($r = 0.49$) 49% en is $.49 < .50$ wat betekent dat de effectgrootte nog net middelmatig is.

De totale scores zowel bij de oude als de nieuwe notatie zijn beide gelijk maar de *Mean Ranks* ($MR = 6.75$) van de oude notatie en de nieuwe ($MR = 6.25$) verschillen van elkaar. De nieuwe notatie heeft meer herhalende waarden in zijn score waardoor hun posities niet doortellen maar de rank verandert in een gemiddelde, daarnaast heeft de oude notatie twee hoge scores waardoor de *rank*

waarde hoog is. Uit de algemene test gemeten op de totaalscore blijkt dat het resultaat statistisch niet significant is waardoor H_a wordt verworpen en H_0 wordt aanvaard.

H_a -Problem solving effectiveness : Het aantal correct uitgevoerde *problem solving tasks* van de nieuwe PGA notatie is hoger dan die van de originele PGA notatie

H_0 -Problem solving effectiveness : Het aantal correct uitgevoerde *problem solving tasks* van de nieuwe PGA notatie is niet hoger dan die van de originele PGA notatie

Tabel 4.6.4 Overzicht resultaten hypothese toets *problem solving tasks* voor *Problem solving effectiveness*

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H_0	H_a
<i>Problem solving effectiveness</i>	<i>problem solving tasks</i>	.093	.186	8 oude notatie	6.63 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.57 nieuwe notatie		

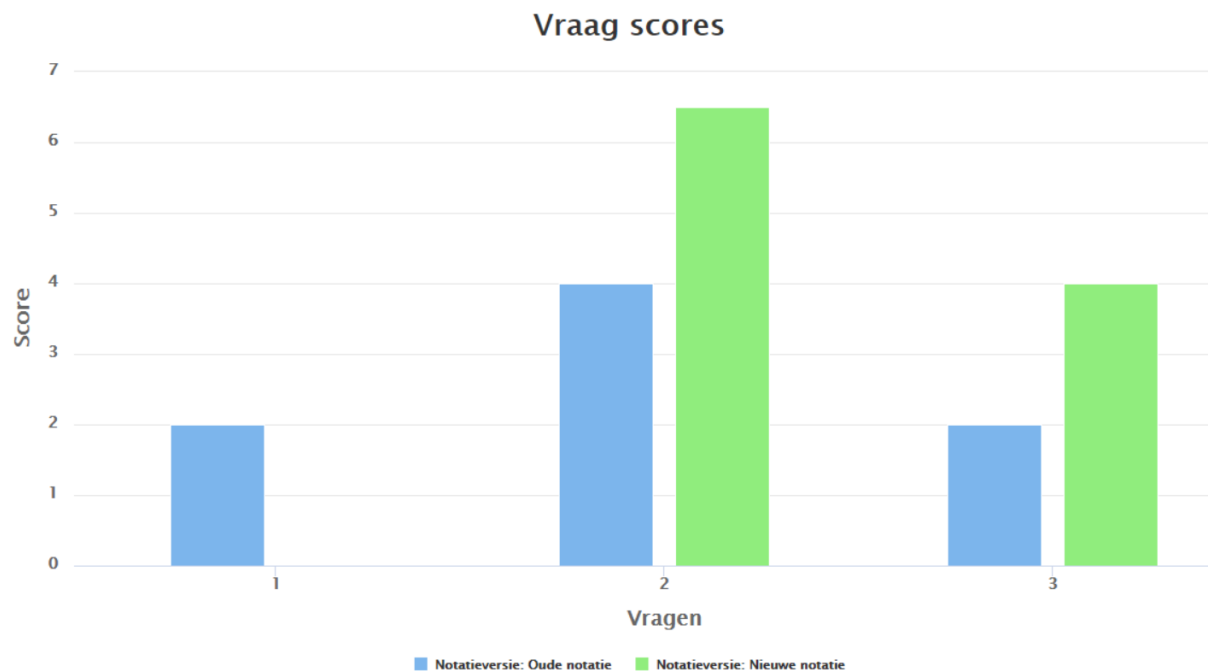
Tabel 4.6.4 toont aan dat het aantal correcte antwoorden in *problem solving tasks* in de nieuwe notatie niet leiden tot een statistische significant resultaat alhoewel deze taak het iets beter lijkt te doen dan de *comprehension questions* ($p = .601$)

Tabel 4.6.5 *problem solving tasks* per vraag voor *Problem solving effectiveness*

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H_0	H_a
<i>Problem solving effectiveness</i>	Vraag 1 (<i>value proposition</i> vervangen door <i>competence</i>)	0.915	.170	8 oude notatie	8.88 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	7.00 nieuwe notatie		
<i>Problem solving effectiveness</i>	Vraag 2 (te hoge baktijd)	0.010	.021	8 oude notatie	5.75 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	10.57 nieuwe notatie		
<i>Problem solving effectiveness</i>	Vraag 3 (Onkosten efficiënter afhandelen)	0.110	.221	8 oude notatie	6.88 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.29 nieuwe notatie		

De tabel uit figuur 4.6.5 toont de uitslag van de Mann Whitney U test per vraag. Vraag 2 (te hoge baktijd) heeft een p -waarde van .010 waaruit we kunnen vaststellen dat de vraag statistisch significant is. De overige vragen worden als niet statistisch significant aangeduid. Kijken we naar de

overige *MRs* van de vragen, dan zien we dat vraag 1 het beste scoort bij de oude notatie. De respondent moest bij deze vraag erachter komen dat de icoon *Competence* de afbeelding had van de *Value proposition*. Het is gebleken dat de algemene test voor de *problem solving effectiveness* geen significant resultaat geeft ondanks het feit dat vraag 2 (PGA-model voor de te hoge baktijd) wel significant is.



Figuur 4.6.6 *problem solving tasks* vragen afgezet tegen scores

We zien in figuur 4.6.6 de vragen 2 en 3, waarbij de nieuwe notatie beter scoort dan de oude de *MRs* zijn > 50% waardoor het waargenomen effect gaat in de richting van de alternatieve hypothese H_a . In vraag 2 wordt aan de respondenten gevraagd om a.d.h.v. het PGA model van een bakkerij het probleem te ontdekken van de te hoge baktijd en in vraag 3 modelleert het PGA-model een willekeurig bedrijf waar de respondenten gevraagd wordt het probleem te ontdekken van de inefficiënte afhandeling van onkosten. Hoewel de vragen 2 en 3 beide hoger scoren in de nieuwe notatie, geeft alleen vraag 2 een significant resultaat ($p = .010$) met $r = .59$ (59%). Een verklaring die hiervoor gegeven kan worden is de context. De onervaren respondenten hebben wat meer 'feeling' met het bakproces dan met financiële processen. We zien dan ook in de oude notatie dat deze vraag hoger scoort dan de andere vragen in de oude notatie, maar het zou dus heel goed kunnen dat de nieuwe notatie hierin het verschil maakt en dat de nieuwe notatie dus idd beter is te volgen.

Ondanks dat de vragen 2 en 3 hoger scoren in de nieuwe notatie dan in de oude notatie waarvan vraag 2 een significante p-waarde heeft, moeten we concluderen dat: H_a wordt verworpen en H_0 wordt aanvaard.

4.7. H-Interpretational efficiency comprehension questions en H-Interpretational efficiency problem solving tasks

H_a-Interpretational efficiency comprehension questions: De responstijd van de *comprehension questions* is korter dan die van de originele PGA notatie

H₀-Interpretational efficiency comprehension questions: De responstijd van de *comprehension questions* is niet korter dan die van de originele PGA notatie

Bij een goed begrijpbare visuele weergave van het PGA model zal de respondent in staat zijn sneller de relevante gebieden te vinden en dus sneller de taak zal uitvoeren wat dan zal resulteren in een kortere responstijd in de nieuwe notatie dan in de oude notatie.

Voor de data preparatie wordt verwezen naar [Bijlage 5](#).

Tabel 4.7.1a Overzicht resultaten hypothese toets *comprehension questions* voor *Interpretational efficiency*

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
Interpretational efficiency comprehension questions	comprehension questions	.577	.423	6 oude notatie	5.76 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.33 nieuwe notatie		

Uit tabel 4.7.1a blijkt dat de algemene responstijd voor de *comprehension questions* een *p*-waarde waarde heeft van .577 met een *MR* = 5.67 voor de oude notatie en een *MR* = 7.33 voor de nieuwe notatie. De algemene test is gedaan op de Totaaltijd variabele, deze heeft een effectgrootte van *r* = .23 (23%) wat gezien kan worden als een kleine effectgrootte. De tabel 4.7.1b toont de resultaten van de losse responstijden per vraag.

Tabel 4.7.1b *comprehension questions* per vraag voor *Interpretational efficiency*

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
Interpretational efficiency comprehension questions	Vraag 1 (Customer goal)	.685	.631	6 oude notatie	6.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.00 nieuwe notatie		
	Vraag 2	.374	.749	6 oude notatie	6.83 oude notatie	Aanvaarden	Verwerpen

<i>Interpretational efficiency comprehension questions</i>	(Valueproposition)			6 nieuwe notatie	6.17 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	Vraag 3 (Internal goal)	.973	.055	6 oude notatie	4.50 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	8.50 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	Vraag 4 (valueStream-Importance)	.869	.262	6 oude notatie	5.33 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.67 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	Vraag 5 (valueStream)	.500	1.000	6 oude notatie	6.50 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.50 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	Vraag 6 (valueStream-Importance-competence)	.685	.631	6 oude notatie	6.00 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.00 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	Vraag 7 (Goals-Valueproposition)	.500	1.000	6 oude notatie	6.50 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.50 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	Vraag 8 (valueStream-Valueproposition)	.739	.522	6 oude notatie	5.83 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	7.17 nieuwe notatie		
<i>Interpretational efficiency comprehension questions</i>	Vraag 9 (valueStreams-Importance-performance)	.626	.749	6 oude notatie	6.17 oude notatie	Aanvaarden	Verwerpen
				6 nieuwe notatie	6.83 nieuwe notatie		
	Vraag 10	.436	.873	6 oude notatie	6.67 oude notatie	Aanvaarden	Verwerpen

<i>Interpretational efficiency comprehension questions</i>	(competence)			6 nieuwe notatie	6.33 nieuwe notatie		
--	--------------	--	--	------------------	---------------------	--	--

In tabel 4.7.1b zien we dat vraag 2 (*valueStream*) met een p -waarde van .374 en een $MR = 6.83$ in de oude notatie en een $MR = 6.17$ in de nieuwe notatie en vraag 10 (*Competence*) ($p = .436$) met een $MR = 6.67$ in de oude en $MR = 6.33$ in de nieuwe notatie het iets beter doen dan de overige elementen, maar hun verschil te klein is voor een significant resultaat. De effectgrootte van deze variabelen zijn resp. $r = .05$ (5%) wat kan worden vertaald naar een kleine effectgrootte en $r = .09$ (9%) wat eveneens een kleine effectgrootte heeft.

Uit het bovenstaande blijkt dat we de hypothese H_a moeten verwerpen en dat H_0 moeten aanvaarden.

H_a -Interpretational efficiency problem solving tasks: De responstijd van de *problem solving tasks* is korter dan die van de originele PGA notatie

Bij een goed begrijpbare visuele weergave van het PGA model zal de respondent in staat zijn sneller de relevante gebieden te vinden en dus sneller de taak zal uitvoeren.

H_0 -Interpretational efficiency problem solving tasks: De responstijd van de *problem solving tasks* is niet korter dan die van de originele PGA notatie

Tabel 4.7.2a Overzicht resultaten hypothese toets *problem solving tasks* voor *Interpretational efficiency*

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H_0	H_a
<i>Interpretational efficiency problem solving tasks</i>	<i>problem solving tasks</i>	.877	.247	8 oude notatie	6.75 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.43 nieuwe notatie		

Uit tabel 4.7.2a blijkt dat de algemene responstijd voor de *problem solving tasks* een p -waarde van .877 heeft met een $MR = 6.75$ voor de oude notatie en een $MR = 9.43$ voor de nieuwe notatie. De algemene test is gedaan op de Totaaltijd variabele, deze heeft een effectgrootte van $r = .30$ (30%) dat gelijk is aan een middelmatig effectgrootte, maar wat wel het omgekeerde effect is van wat men zou verwachten. De tabel 4.7.2b toont de resultaten van de losse responstijden per vraag.

Tabel 4.7.2b *problem solving tasks* per vraag voor *Interpretational efficiency*

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
Interpretational efficiency problem solving tasks	Vraag 1 (value proposition vervangen door competence)	.851	.298	8 oude notatie	6.88 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.29 nieuwe notatie		
Interpretational efficiency problem solving tasks	Vraag 2 (te hoge baktijd)	.851	.298	8 oude notatie	6.88 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.29 nieuwe notatie		
Interpretational efficiency problem solving tasks	Vraag 3 (Onkosten efficiënter afhandelen)	.823	.355	8 oude notatie	7.00 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.14 nieuwe notatie		

Tabel 4.7.2b toont de uitslagen van de Mann-Whitney toets per vraag. Het wordt al snel duidelijk dat de oude notatie een kortere responstijd laat zien dan de nieuwe en dat de p -waardes allen ruim boven de significantie waarde liggen ($p > .05$). Daarom zal de hypothese H_a worden verworpen en H₀ aanvaard.

4.8. H-Perceived ease of use

H_a-Perceived ease of use : Het waargenomen gebruiksgemak van de nieuwe PGA notatie is hoger dan die van de originele PGA notatie.

H₀-Perceived ease of use : Het waargenomen gebruiksgemak van de nieuwe PGA notatie is niet hoger dan die van de originele PGA notatie.

Na afloop van het experiment werd aan de hand van een vragenlijst (zie figuur 4.6.1) de respondenten gevraagd naar het gebruiksgemak van de notatie. Het resultaat van deze hypothese zal de bovenstaande hypothesen ondersteunen.

Voor de datapreparatie wordt verwezen naar [Bijlage 5](#).

Vraag over gebruik van de modelleertaal

🔴 Dit is een verplichte vraag
🔴 Vul alle onderdelen in

	1 - Zeker niet mee eens	2	3	4 - Niet mee eens of oneens	5	6	7 - Zeker mee eens
Het was voor mij gemakkelijk te begrijpen wat het model wilde modelleren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Het gebruik van het model was vaak frustrerend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Over het algemeen was het gebruik van het model eenvoudig te gebruiken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Het leren lezen van het model was makkelijk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figuur 4.8.1 *questionnaire* voor de *perceived ease of use*

Tabel 4.8.2a Overzicht resultaten hypothese toets *perceived ease of use*

Variabele	Taak	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
<i>perceived ease of use</i>	<i>perceived ease of use</i>	.016	.032	11 oude notatie	7.36 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	12.86 nieuwe notatie		

De nieuwe notatie uit tabel 4.8.2a laat een significant resultaat zien dat de score hoger is van de nieuwe notatie dan van de oude notatie. De $MR = 7.36$ voor de oude en $MR = 12.86$ voor de nieuwe notatie en heeft een p -waarde van .016. Tabel 4.8.2b toont de individuele scores van de vragen.

Tabel 4.8.2b *perceived ease of use* per vraag

Variabele	Vraag	p-waarde (1-tailed)	p-waarde (2-tailed)	Aantal cases	Mean Rank	H ₀	H _a
<i>perceived ease of use</i>	Vraag 1 (begrijpen wat model wil modelleren)	.556	.888	11 oude notatie	9.64 oude notatie	Aanvaarden	Verwerpen
				7 nieuwe notatie	9.29 nieuwe notatie		
<i>perceived ease of use</i>	Vraag 2 (frustratie)	.024	.048	11 oude notatie	7.68 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	12.36 nieuwe notatie		
	Vraag 3	.012	.024	11 oude notatie	7.32 oude notatie	Verwerpen	Aanvaarden

<i>perceived ease of use</i>	(eenvoudig te gebruiken)			7 nieuwe notatie	12.93 nieuwe notatie		
<i>perceived ease of use</i>	Vraag 4 (leren lezen van model)	.039	.079	11 oude notatie	7.82 oude notatie	Verwerpen	Aanvaarden
				7 nieuwe notatie	12.14 nieuwe notatie		

Uit tabel 4.8.2b is af te lezen dat de respondenten in die de nieuwe notatie hadden het lastiger vonden te begrijpen wat er gemodelleerd moest worden dan respondenten die de oude notatie hadden (zie vraag 1). Het is zelfs zo dat deze vraag de hoogste score heeft in de oude notatie t.o.v. de overige vragen in de oude notatie. De uitslag voor deze vraag is niet significant ($p = .556$) dus kunnen we hier verder geen uitspraken over doen. Wel kunnen we zien dat, doordat het resultaat niet significant is, het verschil in MR s tussen de oude en nieuwe notatie ($MR_{oud} - MR_{nieuw} = .35$) klein is.

De frustratie (zie vraag 2) was vooral aanwezig bij de respondenten die gebruik maakte van de oude notatie. Merk op dat vraag 2 'negatief' is gesteld en dat de codering voor deze vraag is omgenummerd (zie [Bijlage 5](#)). Kijken we naar de p -waarde van vraag 2 dan is deze .024, een $MR = 7.68$ voor de oude notatie en $MR = 12.36$ voor de nieuwe notatie. De effectgrootte is $r = .47$ (47%) wat een middelmatige effectgrootte heeft. De effectgrootte van vraag 3 en 4 zijn resp. $r = .53$ (53%) een grote effectgrootte heeft en $r = .41$ (41%) wat gelijk is aan een middelmatig effectgrootte. Uit deze test blijkt dat respondenten die de nieuwe notatie voor zich hadden minder moeite hadden het model te gebruiken dan de respondenten die de oude notatie hadden. Voor vraag 3 (de mate van gebruiksgemak) heeft dit vooral te maken dat wanneer het model makkelijk te lezen is (zie vraag 4) het een overzichtelijk model is en daardoor de relevante gebieden betrekking hebbend op een vraag eenvoudig zijn te vinden. Men hoefde alleen maar verticaal te scrollen, maar dat niet alleen, doordat de *valueStream* elementen in de nieuwe notatie voorzien zijn van pijlen fungeren deze a.h.w. als richting aangevers waardoor het gebruiksgemak nog meer versterkt wordt.

We kunnen stellen dat H_0 kan worden verworpen en H_a kan worden aanvaard.

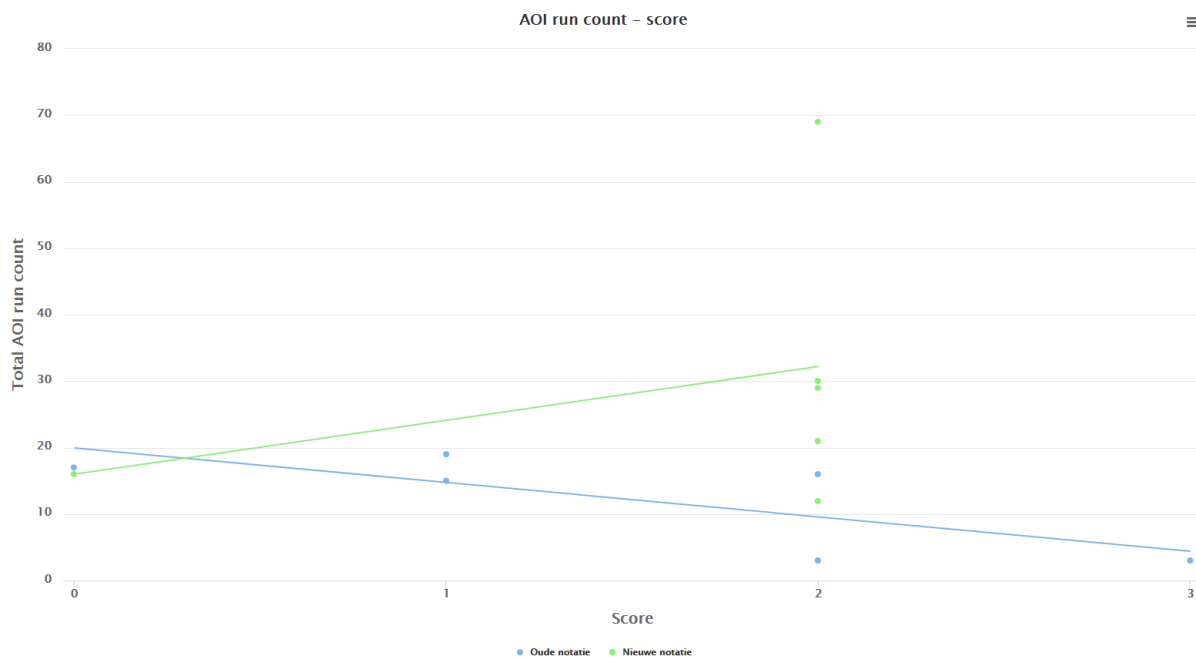
4.9. AOI run count en AOI fixatie duur - Score

Deze paragraaf staat los van de hypotheses gedefinieerd in hoofdstuk 3. Het kan gezien worden als een extra paragraaf die de resultaten uit de hypotheses AOI fixatie *run count* en AOI fixatie duur afzet tegen de resultaten van de effectiviteit om te kijken of er een verband bestaat.

Aangezien de AOI fixatie duur voor de *comprehension questions* geen significant resultaat geeft, wordt dit niet in deze paragraaf meegenomen.

We verwachten positieve lineaire verbanden te vinden tussen de variabelen, immers hoe hoger de AOI fixatie *run count* en duur hoe hoger de score zou moeten zijn. Wanneer de visualisatie in de nieuwe notatie duidelijk te volgen en te begrijpen is, zal een respondent bij een vraag de bijbehorende relevante gebieden beter kunnen vinden in deze notatie dan bij de oude. De verwachting is dat er bij de oude notatie ook veel meer niet-relevante gebieden worden bezocht, dus het aantal keren dat zo'n relevant gebied wordt bezocht zal hoger liggen in de nieuwe notatie

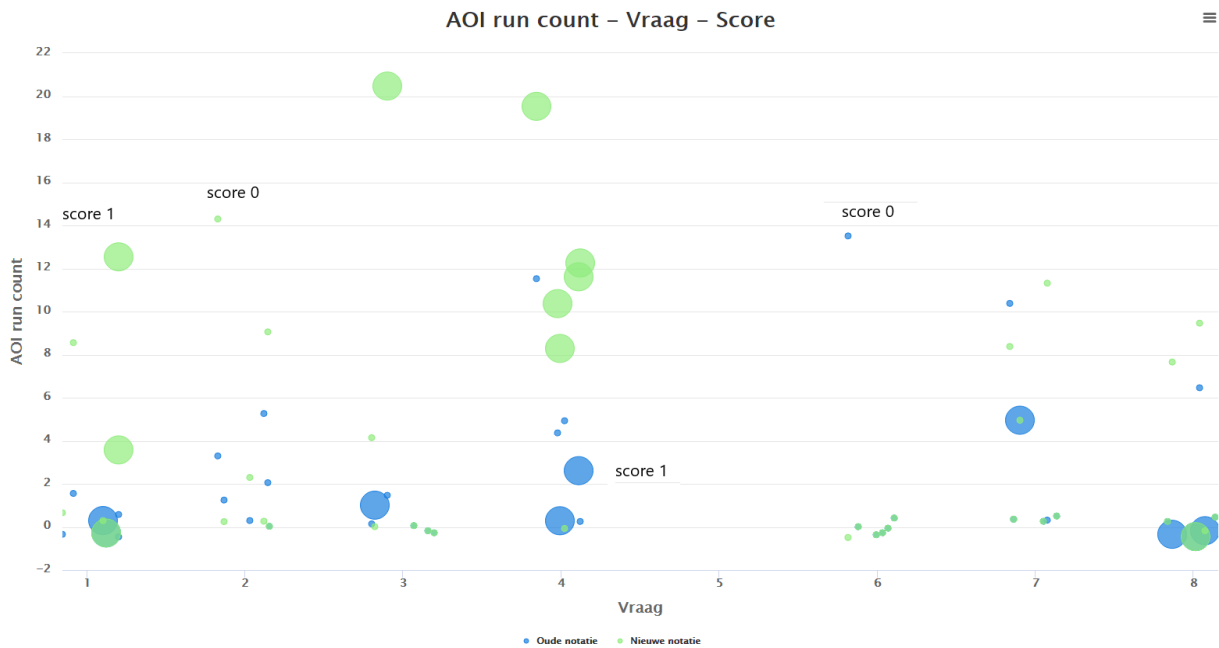
dan bij de oude notatie. Voor de AOI fixatie duur gaat het om de tijd die wordt besteed aan een relevant gebied. Een respondent zal in de oude notatie een minder lange fixatie duur hebben in een relevant gebied (AOI) dan in de nieuwe notatie omdat hij vanwege een minder duidelijke visualisatie sneller gaat zoeken naar andere gebieden om te kijken of hij/zij daar (wel) herkenbare punten kan vinden (Borkin, 2016; Bera, 2019). Het verwachte effect op de effectiviteit houdt in dat de score in de nieuwe notatie hoger is dan in de oude notatie. Hieruit volgt dat bij een hoge AOI *run count* en fixatie duur de visualisatie beter wordt begrepen en daardoor een hogere score zal hebben wat naar een positief lineaire verband wijst.



Figuur 4.9.1a *comprehension questions* AOI *run count* afgezet tegen scores

Figuur 4.9.1a visualiseert de relatie tussen AOI *run count* en de score van de comprehension questions op respondent niveau. Elk datapunt stelt een respondent voor die een totaal aan AOI *run count* heeft behaald met een totaal score over alle vragen. Men verwacht een hogere score in de AOI *run count* bij de nieuwe notatie dan bij de oude notatie, simpelweg omdat deze beter te begrijpen zou zijn. Hieruit volgt dat wanneer een visualisatie beter begrepen wordt dat de score in de antwoorden ook hoger is. We zien een stijgende lijn in de nieuwe notatie, maar wat we ook zien is dat de oude notatie een licht negatief verband toont tussen de twee variabelen.

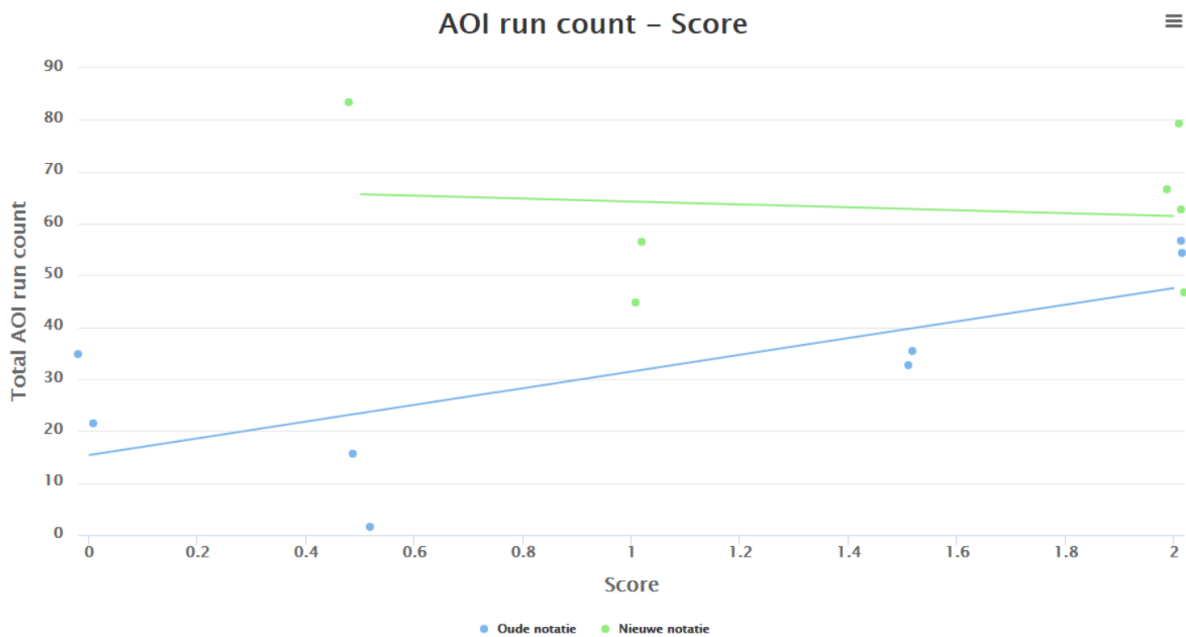
Dus in dit geval hoe lager de AOI *run count*, hoe hoger de score. In de oude notatie lijkt het dat wanneer een respondent een lage AOI *run count* heeft hij/zij een hogere score behaald. Figuur 4.9.1b *zoomt* dieper in op dit resultaat.



Figuur 4.9.1b *comprehension questions* AOI run count – Vraag - scores

Figuur 4.9.1b toont op de x-as de vragen. Deze staan op chronologische volgorde in tijd, m.a.w. vraag 1 is in tijd eerder aan de beurt dan vraag 8. Op de y-as staan de AOI *run count* van een respondent en verder zien we dat de groene kleur de nieuwe notatie vertegenwoordigt en blauw de oude. De grootte van de cirkels geeft de hoogte van de behaalde score aan, nl. de waarde 1 of 0, resp. grote of kleine cirkel per respondent. Zoals al eerder in paragraaf 4.5 werd vermeld willen we nogmaals aangeven dat vraag 5 (*valueStream*) niet werd gevonden door de *eye-tracking* software en dat voor de vragen 9 en 10, betrekking hebbend op de elementen *valueStreams-Importance-performance* en *competence*, geen relevante gebieden werden gevonden door de respondenten.

De vraag die er uitspringt is vraag 4 (*valueStream-importance*), nl. alle respondenten hebben de relevante AOI gebieden gevonden. We zien dat de meest goede antwoorden voorkomen in de nieuwe notatie, bovendien tellen de respondenten in de nieuwe notatie bij deze vraag ook meer *run counts* dan de in de oude. Ditzelfde geldt ook voor vraag 1 (*customer goal*) al hebben we hier te maken met maar 3 respondenten die de relevante gebieden binnengaan met goede antwoorden in de nieuwe notatie tegen 1 respondent met één goed antwoord van de oude notatie die een relevant gebied heeft gevonden waarbij de *run count* in de oude notatie vele male kleiner is dan bij de nieuwe notatie. Bij de vragen 7 en 8 (*valueStream-Importance-competence*, *goals-Valueproposition*, *valueStream-Valueproposition*) zijn de scores in de nieuwe notatie lager dan bij de oude. De hoge scores uit figuur 4.9.1a voor de oude notatie, worden vooral bepaald door de vragen 7 en 8 (*Goals-Valueproposition*, *valueStream-Valueproposition*). Opvallend is dat er een afname van het vinden van relevante gebieden plaatsvindt naarmate men verder komt met de vragen (of in de tijd). We zien ook dat de respondenten van de nieuwe notatie vanaf vraag 6 aanzienlijk slechter scoren dan de respondenten die de oude notatie hebben. Dit verklaart de dalende lijn uit figuur 4.9.1a en de stijgende lijn van de oude notatie, deze wordt voor een groot deel bepaald door de score van vraag 7.

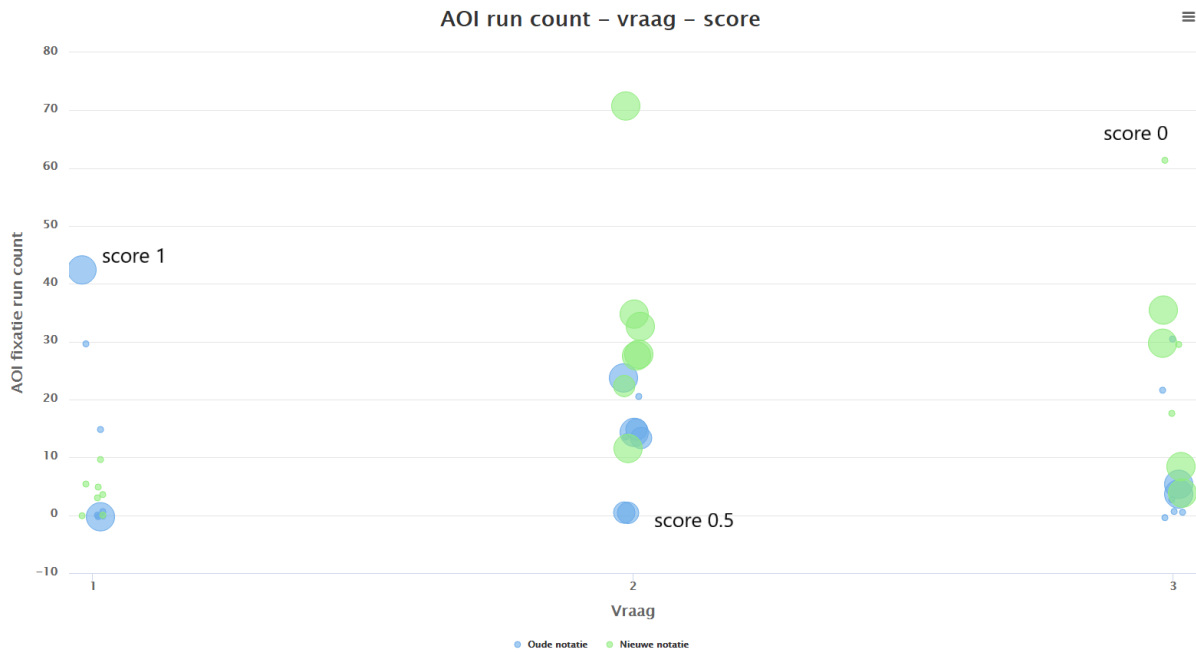


Figuur 4.9.2a *problem solving tasks* AOI run count afgezet tegen scores

Figuur 4.9.2a visualiseert de relatie tussen AOI run count en de score van de *problem solving tasks* per respondent. We zien vooral in de oude notatie een duidelijk positief lineair verband terug. De nieuwe notatie laat een zwak (zelfs negatieve) lineaire verband zien. Dit komt vooral door één respondent die een hoge AOI fixatie run count heeft over alle vragen maar tevens ook het laagst scoort over alle vragen. De twee notatie versies tonen hierdoor een contrasterend beeld.

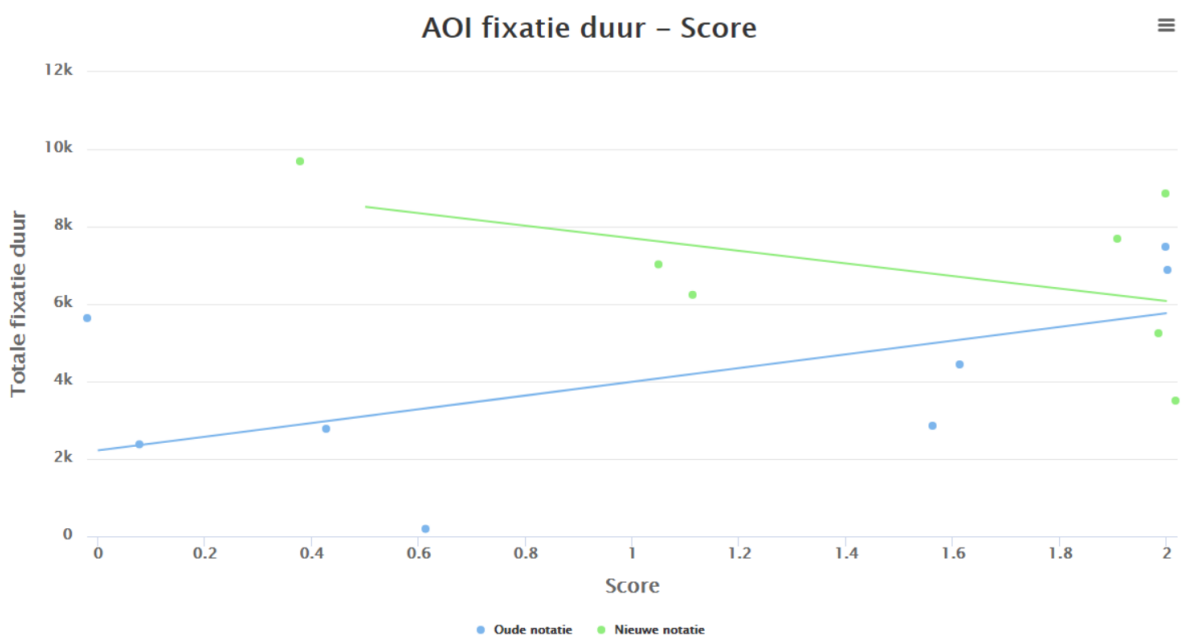
Figuur 4.9.2b toont net als in figuur 4.9.1b op de x-as de vragen. Deze staan op chronologische volgorde in tijd. M.a.w. vraag 1 is in tijd eerder aan de beurt dan vraag 3. Op de y-as staan de AOI run count van de respondenten en verder zien we dat de blauwe kleur de oude notatie vertegenwoordigt en groen de nieuwe. De grootte van de cirkels geeft de hoogte van de scores weer, dus hoe groter de cirkel hoe hoger de score.

Van paragraaf 4.5 weten we dat de hoogste scores voor beide notaties te vinden zijn bij de vragen 2 (te hoge baktijd) en 3 (Onkosten efficiënter afhandelen) waarbij in beide gevallen de nieuwe notatie hoger scoort dan de oude notatie. Wanneer we naar de AOI fixatie run counts kijken per vraag in figuur 4.9.2b valt op dat vraag 2 in de nieuwe notatie bijzonder goed is vertegenwoordigd met de hoge run counts dat op de voet wordt gevolgd door vraag 3. Bij vraag 1 (*value proposition* vervangen door *competence*) zien we dat de nieuwe notatie niets (0) scoort en dat de oude notatie een 2 scoort, deze versie toont ook hoge run counts, terwijl de nieuwe notatie lage run counts laat zien. De score (2) bij de oude notatie in figuur 4.9.2a is daarom ook afkomstig van vraag 1.



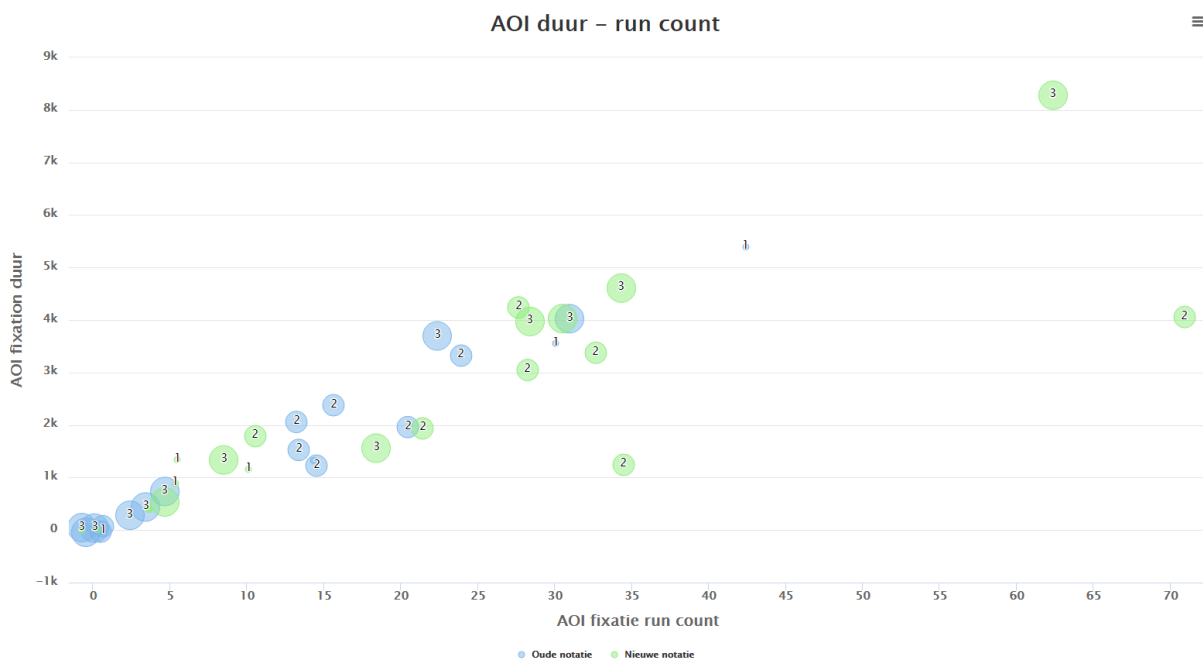
Figuur 4.9.2b *problem solving tasks* AOI run count – Vraag – scores

De AOI fixatie duur voor de *problem solving tasks* had in paragraaf 4.5 een significant resultaat ($p = 0.0245$). In deze paragraaf wordt er gekeken of er een verband bestaat tussen de AOI fixatie duur en de behaalde score van de problem-solving questions. Figuur 4.9.3 geeft duidelijk de lineaire verbanden weer waarbij de negatieve lijn ook hier te zien is in de nieuwe notatie en de positieve lijn in de oude notatie.



Figuur 4.9.3 *problem solving tasks* AOI fixatie duur afgezet tegen scores

Tot slot is in onderstaand figuur 4.9.4 te zien dat de AOI fixatie duur en *run count* onderling een positief lineair verband tonen wat betekent dat beide metingen in dit experiment dezelfde informatie bevatten. In de cirkels staan de vraagnummers vermeld. Hier en daar zien we uitschieters in de nieuwe notatie bij de vragen 2 (te hoge baktijd) en 3 (Onkosten efficiënter afhandelen) in de vorm van een hoge AOI fixatie duur of *run count*. Hier zou men uit kunnen concluderen dat beide metingen (fixatieduur en *run count*) vanwege de sterke correlatie dezelfde informatie weergeven en dat de afwijkingen het meest zichtbaar zijn in de AOI fixatie *run count*. Men zou dus kunnen stellen dat de AOI fixatie *run count* in deze analyse van deze paragraaf ons kan voorzien van de meest relevante informatie in relatie tot de behaalde scores.



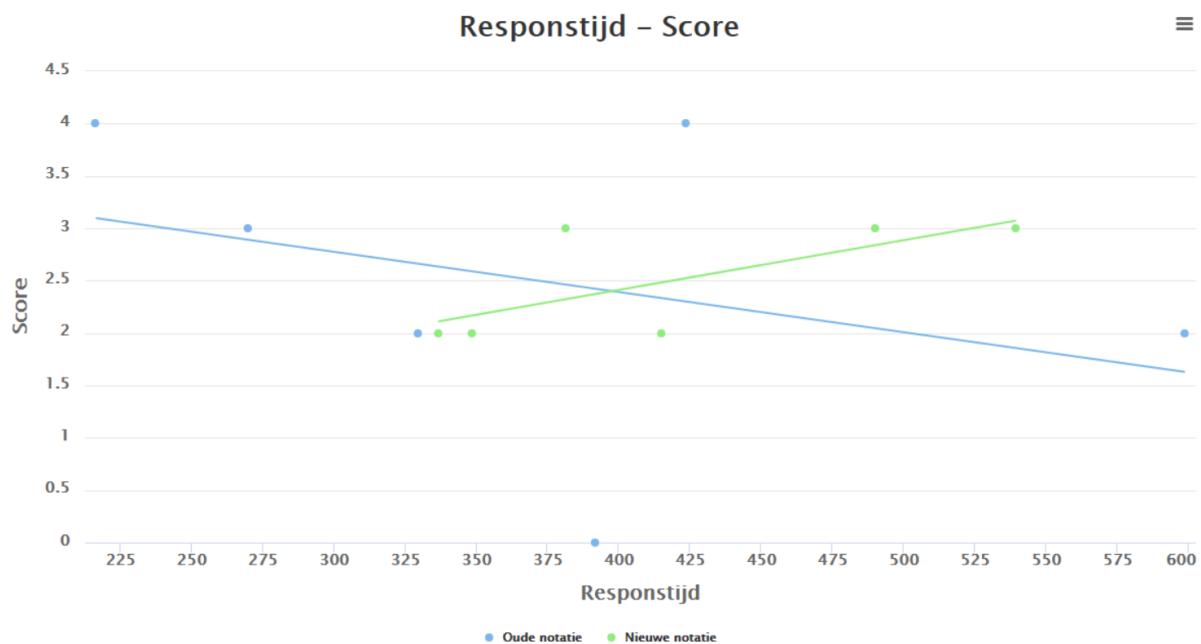
Figuur 4.9.4 relatie tussen AOI *run count* en AOI fixatieduur

4.10. Responsetijd – Score

In paragraaf 4.7 (*Interpretational efficiency*) hebben we moeten constateren dat de hypothese toetsen geen significant resultaat gaven. Deze testen bestonden uit een algemene test waarbij een niet-parametrische independent-sample test werd gebruikt om de twee groepen, nieuwe en oude notatie, te toetsen of de nieuwe notatie een kortere responstijd gaf dan de oude notatie. De afhankelijke variabele die hiervoor werd gebruikt was de geaggregeerde responstijd van de vragen per respondent per notatie. Verder werd een detail test gedaan op de vragen per taak, *comprehension questions* en *problem solving tasks*. Bij de *comprehension questions* scoorde *valueStream* en *competence* beter t.o.v. de overige elementen. Bij de *problem solving tasks* waren er geen opdrachten/vragen die er uitsprongen. Kortom, op basis van de responstijd kan men geen significante verschillen identificeren.

In deze paragraaf worden de responstijden van de *comprehension questions* en *problem solving tasks* afgezet tegen de behaalde scores en gekeken of er een negatief lineaire verband te vinden is.

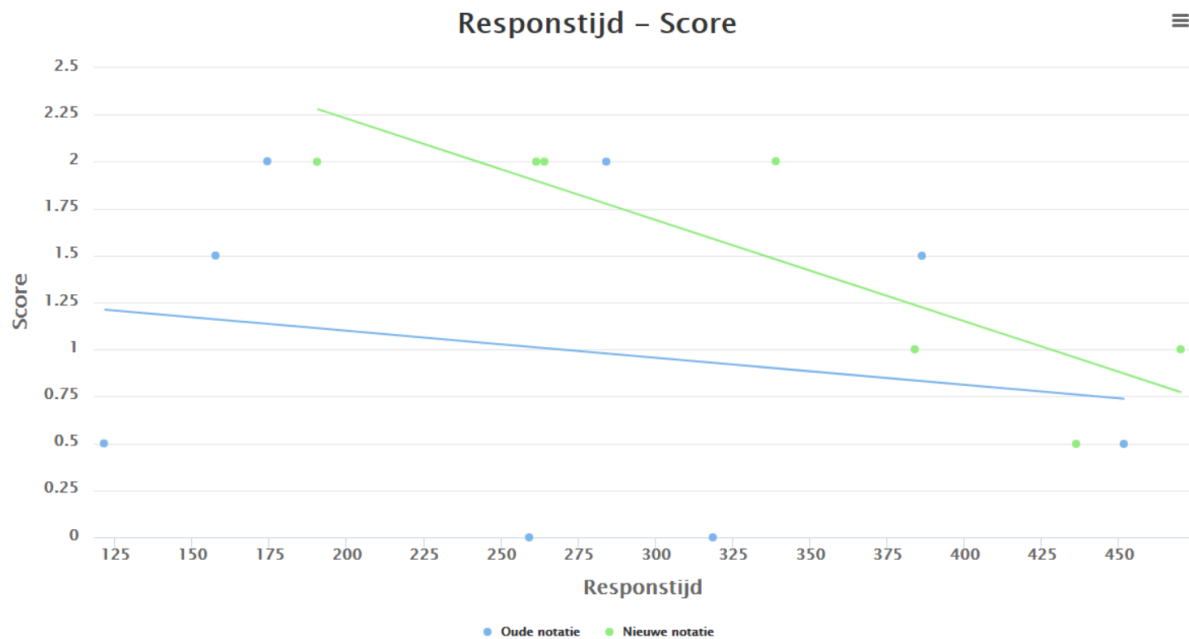
Dit wordt onderzocht omdat we er vanuit gaan dat een korte responstijd aangeeft dat de visualisatie duidelijk en begrijpelijk is waardoor de respondent minder tijd nodig heeft om te antwoorden. Wanneer de respondent minder tijd nodig heeft om te antwoorden omdat hij/zij de visualisatie begrijpt zal dan ook de score hoger zijn naarmate de responstijd korter is. In dat geval ontstaat er een negatief lineaire verband.



Figuur 4.10.1 *comprehension questions* responstijd afgezet tegen scores

Figuur 4.10.1 toont een spreidingsdiagram met de responstijd en score van de *comprehension questions*.

We zien een korte positieve lijn van de nieuwe notatie en een lange dalende lijn van de oude notatie. Het is opvallend dat naar mate de responstijd oploopt de score in de oude notatie omlaag gaat. Het verwachte negatieve lineaire verband is enkel en alleen maar te zien in de oude notatie waarbij we het tegenovergestelde zien in de nieuwe notatie.



Figuur 4.10.2 *problem solving tasks* responstijd afgezet tegen scores

Figuur 4.10.2 toont het verband tussen Responstijd en Score van de *problem solving tasks*. Zowel de oude als nieuwe notatie gaan in score omlaag naar mate de responstijd oploopt. Zowel in de oude als in de nieuwe notatie zien we een negatief lineaire verband, waarbij de nieuwe notatie, zo lijkt het een sterker verband lijkt te hebben met de hoogte van de score en de responstijd. We zien ook dat naarmate de responstijd oploopt de verschillen tussen de notaties steeds kleiner wordt.

5. Discussie, conclusies en aanbevelingen

Hier trekt u conclusies uit het onderzoek. Wat betekenen de resultaten en in hoeverre zijn ze vergelijkbaar met resultaten van eerder onderzoek. Hoe kunt u de eigen resultaten positioneren t.o.v. de literatuur en wat kan men er dus echt van leren. Deze discussie is een substantieel deel van het rapport en vergt minimaal één, maar bij voorkeur twee of drie pagina's.

U kunt de volgende kopjes gebruiken om de discussie te structureren.

5.1. Discussie

Doel van dit onderzoek is om de resultaten uit het onderzoek van Roelens & Bork (2020), waarin de oude en nieuwe PGA-notatie worden geëvalueerd op intuïtiviteit, middels statistische methoden te evalueren. Het onderzoek van Roelens & Bork (2020) is immers een kwalitatief onderzoek waaraan studenten als respondenten hebben deelgenomen. De resultaten uit dat onderzoek, hiermee bedoelen we de nieuw voorgestelde notatie als zijnde een verbetering op de oude notatie, zijn met een combinatie van creativiteit en interpretatie door de onderzoekers tot stand gekomen. Doel van dit onderzoek is om de resultaten uit het kwalitatief onderzoek van Roelens & Bork (2020) statistisch te evalueren.

5.1.1. Resultaten

In deze paragraaf volgt een samenvatting van de resultaten. Het beschrijft wat uit de algemene tests naar voren is gekomen en de bijzonder-/opmerkelijkheden. Verder wordt er geprobeerd om a.d.h.v. de testresultaten een verklaring te vinden vanuit de gevonden literatuur. De testresultaten worden in volgorde van uitvoering zoals deze voorkomt in hoofdstuk 4 besproken.

Tabel 4.10.1 Overzicht eindresultaat hypothese toetsen

Variabele	Taak	H _a	Significant gevonden elementen/vragen
Algehele fixatie aantal	<i>comprehension questions</i>	Verwerpen	n.v.t
Algehele fixatie duur		Verwerpen	
AOI fixatie <i>run count</i>		Aanvaarden	Vraag 1 (<i>Customer goal</i>)
AOI fixatie duur		Verwerpen	n.v.t.
<i>Comprehension questions effectiveness</i>		Verwerpen	Vraag 4 (<i>valueStream-Importance</i>)
<i>Interpretational efficiency comprehension questions</i>		Verwerpen	n.v.t

Algehele fixatie aantal	<i>problem solving tasks</i>	Verwerpen	n.v.t
Algehele fixatie duur		Verwerpen	
AOI fixatie <i>run count</i>		Aanvaarden	Vraag 2(bakkerij model)
			Vraag 3 (onkosten afhandelen)
AOI fixatie duur		Aanvaarden	Vraag 3 (onkosten afhandelen)
<i>Problem solving effectiveness</i>		Verwerpen	Vraag 2 (bakkerij model)
<i>Interpretational efficiency</i> <i>problem solving tasks</i>		Verwerpen	n.v.t
<i>perceived ease of use</i>	<i>perceived ease of use</i>	Aanvaarden	Vraag 2(frustrerend)
			Vraag 3(gebruik van het model)
			Vraag 4(leren lezen van het model)

Paragraaf 4.4 Hypotheses H-fixatie duur en H-fixatie aantal

Het blijkt dat de nieuwe notatie een hogere totale fixatieduur en -aantal heeft dan de oude notatie. Dit resultaat is ook grotendeels terug te zien bij de metingen per vraag. Volgens Borkin et al. (2016) zullen minder duidelijke visualisaties meer tijd nodig hebben om te zoeken naar herkenbare punten dan de duidelijke visualisaties. De fixaties zouden hoger moeten zijn bij de onduidelijke visualisaties dan bij duidelijke. In dit onderzoek zien we het omgekeerde resultaat. De resultaten zijn dan ook niet significant te noemen. Een verklaring zou kunnen zijn dat het omgekeerde effect aangeeft dat respondenten juist wel duidelijke aanknopingspunten kunnen vinden tussen de visualisatie en het langetermijngeheugen. En doordat er bij de oude notatie geen herkenningpunten gevonden kunnen worden juist eerder wordt opgegeven waardoor fixatieduur en aantal korter zijn.

Paragraaf 4.5 Hypotheses H-AOI fixatie duur en H-AOI run count.

De algemene hypothese toetsen leidt tot verschillende uitkomsten, nl. de *comprehension questions* geeft een niet significant resultaat voor de AOI fixatie duur maar wel voor AOI fixatie *run count* en voor de *problem solving tasks* geldt voor zowel AOI fixatie duur als voor AOI fixatie *run count* een significant resultaat.

Het blijkt over het algemeen dat de nieuwe notatie een hogere AOI fixatie *run count* en duur heeft dan de oude notatie ondanks dat de duur in de *comprehension questions* niet significant is.

Bij de *comprehension questions* (AOI fixatie *run count*) zien we vooral dat de elementen *Customer goal* en de combinatie *valueStream - Importance* goed scoren waarbij het element *Customer goal* een significant resultaat geeft ($p = .30$).

Volgens Bera et al. (2019) is bij een visualisatie die goed begrepen wordt de AOI fixatieduur en *run count* hoger dan bij een visualisatie die niet of minder goed wordt begrepen. Dit is in lijn met wat we zouden verwachten in onze resultaten. Een verklaring waarom de elementen *valueStream*, *Importance* en *Customer goal* beter in de nieuwe dan in de oude scoren kan volgens Moody (2009) te maken hebben met een aantal van de negen principes die worden genoemd in "*The 'Physics' of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering*", nl. *Semantic transparency*. Het doel is om de cognitieve belasting te reduceren door gebruik te maken van zgn. ingebouwde geheugensteun steuntjes. Hiermee wordt bedoeld dat wanneer alledaagse objecten worden gebruikt men daar eenvoudig een betekenis aan kan geven. Zo verdient het de voorkeur om functioneel vergelijkbare objecten te gebruiken. Bijv. een vuilnisbak icoontje kan een verwijder actie betekenen. In het licht van het *valueStream* element in de vorm van een pijl kan er een duidelijke en functionele betekenis aan gegeven worden. Hetzelfde geldt voor het *Importance* element. De uitroep tekens worden in het dagelijks leven vaak gebruikt als teken om op te letten of om inderdaad aan te geven dat iets belangrijk is. Het volgende principe dat een verklaring kan geven is *semiotic clarity* een symbool moet een één-op-één mapping hebben met zijn semantische constructie. Zo zou verklaard kunnen worden dat het symbool van de pijl *valueStream* een logische plaats heeft in het model dat a.h.w. een *flow* of *stream* aangeeft van het ene element naar het andere en dat de uitroep tekens aangeven dat een bepaalde *stream* de mate van belangrijkheid aangeeft. De *Customer goal* element is een representatie van een Dartbord en twee mensfiguren die elkaar een schouderklopje geven. *Semantic transparency* speelt een belangrijke rol bij het zien van een Dartbord dat een sterke associatie heeft met een doel. Verder is het volgens Borkin et al. (2016) zo dat visualisaties met *Human Recognizable Objects (HRO)* beter herkend en begrepen worden dan visualisaties waarin deze niet of nauwelijks aanwezig zijn. *Competence* scoort in de oude notatie beter dan in de nieuwe. *Semantic transparency* zou ook hier weer een belangrijke rol kunnen spelen, nl. dat een podium voor eerste t/m derde plaats zekere mate van prestatie symboliseert en dat een tekening met een lampje juist meer op bijv. vindingrijkheid, innovatie of ideeën wijst. In allerlei

boeken (o.a. stripboeken) wordt dit symbool aangehaald. Het zou daarom ook logisch zijn dat men in het langetermijngeheugen eerder een associatie vindt met idee, vindingrijkheid of innovatie dan competentie. Echter hoeft het niet zo te zijn dat de grafische afbeelding van het icoon niet duidelijk is. Mendling & Recker (2008) geven aan dat wanneer een grafische afbeelding en/of het bijbehorend tekstlabel op meerdere manieren geïnterpreteerd kan worden het begrijpen van het model moeilijk wordt. In dit geval geloven we dat het *competence* element van de oude notatie een veel duidelijkere grafische afbeelding heeft dan in de nieuwe notatie, maar dat de tekst wellicht niet helemaal aansluit.

De *problem solving tasks* gaat om het begrijpen van een probleem en beredeneren wat de oorzaak kan zijn waardoor het probleem is ontstaan. De manier hoe het model is opgebouwd/geconstrueerd of beter gezegd georganiseerd kan volgens Bera et al. (2019) de zgn. *Domain understanding* vergroten of zelfs verminderen. Vergelijken we de oude notatie met de nieuwe notatie dan geeft de nieuwe notatie een duidelijke flow weer door de *valueStream* elementen. Deze bestaan uit een pijl icoon waardoor het duidelijk wordt dat het model van onder naar boven gelezen moet worden. We gaan er dan ook vanuit dat dit een verklaring kan zijn voor de hogere AOI fixatieduur en *run count*. Vraag 1 scoort in de nieuwe notatie hoger dan in de oude notatie. Hier moest de grafische afbeelding voor *value proposition* vervangen worden door de *competence* afbeelding. *Value proposition* scoort in de *comprehension questions* ook niet hoog. Dit zou erop kunnen wijzen dat *value proposition* geen echte verbetering geeft in de nieuwe notatie en dat het feit dat de AOI fixatieduur en *run count* in de nieuwe notatie hoger scoren meer zegt over de constructie en duidelijkheid van de overige elementen of dat de verwarring die aanwezig is in beide notaties is ontstaan door onjuist gebruik van tekstlabels. Eén respondent die de oude notatie voor zich had gaf het volgende antwoord: "element B bevat 'uitmuntendheid' en dit is een vaag en relatief begrip". Men zou hieruit kunnen afleiden dat het label bij deze vraag voor verwarring en dus voor een lage score heeft gezorgd in de notatie. De vragen 2 (Te hoge baktijd) en 3 (Efficiënt afhandelen van onkosten) hebben beide een significant resultaat. Wat duidelijk blijkt is dat respondenten van de nieuwe notatie de *valueStream* en *Importance* gebruiken om te komen tot een gedegen antwoord: "De iconen F D B en A stellen het kritieke pad voor. Deze iconen zijn verantwoordelijk voor de hoge baktijd." en "Ik heb het idee dat het komt door de onderdelen die in het rood staan. Hierbij vooral het voorverwarmen en het bakken zelf". In de oude notatie wordt het kritieke pad niet genoemd. Respondenten met de nieuwe notatie waren in staat de knelpunten aan te geven wat niet geldt voor de respondenten met de oude notatie. Het lijkt erop dat er meer effort wordt gestoken door de respondenten die de nieuwe notatie voor zich hebben, om een goed antwoord te geven voor het probleem. Wanneer men goed kan beredeneren over een bepaald probleem kan dit wijzen op het feit dat de cognitieve belasting niet te hoog is. Volgens Moody (2009) wordt de *cognitive load* omlaag gehaald door de *Semantic transparency*.

Paragraaf 4.6 Hypotheses H-*Comprehension questions effectiveness* en H-*Problem solving effectiveness*

De algemene tests voor de *comprehension questions* en *problem solving tasks*, geven geen significante resultaten. Wanneer ingezoomd wordt op de afzonderlijke vragen constateren we bij de *comprehension questions* een significant resultaat bij vraag 4 wat betrekking heeft op de elementen *valueStream-Importance* ($p = .046$).

Uit algemene tests voor de *comprehension questions* wordt alleen maar bevestigd dat *valueStream* en *Importance* een duidelijke verbetering geven op de nieuwe notatie, daarnaast lijkt het element *customer goal* het op twee na beste element te zijn dat iets beter scoort in de nieuwe notatie. De verschillen tussen beide prestaties van deze elementen zou verklaard kunnen worden doordat

“Verlagen financiële druk” niet zo direct gekoppeld is aan de grafische afbeelding van twee mensfiguren die elkaar een handdruk geven. Dit zou wat verwarring kunnen geven, aan de andere kant zien we dat dit icoon in de nieuwe notatie het beter doet dan in de oude vanwege het feit dat de twee mensfiguren duidelijker zijn dan de hand afbeeldingen in de oude notatie. Een belangrijk punt om mee te nemen is dat men door het feit dat er op hetzelfde hiërarchisch niveau drie icoontjes staan met een Dartbord, zullen deze dan ook logischerwijs gezien worden als doelen en wel drie verschillende soorten doelen. Wanneer dan in vraag 1 gevraagd wordt welk klantdoel er wordt gesteld zijn er twee respondenten die ‘Geen idee’ hebben ingevuld bij de oude notatie, de overige respondenten (i.e. oude en nieuwe notatie) hebben allemaal één van de drie doelen als antwoord gegeven. Hieruit kunnen we concluderen dat het Dartbord afbeelding een sterke associatie heeft met een ‘doel’. Het feit dat in de oude notatie twee respondenten hebben gekozen voor ‘Geen idee’ zou te maken kunnen hebben volgens Mendling & Recker (2008) met het feit dat de iconen in de oude notatie minder goed van elkaar onderscheiden kunnen worden dan in de nieuwe notatie. Hoe beter iconen van elkaar onderscheiden kunnen worden hoe minder dit kan leiden tot misinterpretaties van een model. We menen dan ook dat door het gebruik van dezelfde kleuren, in de oude notatie, voor *importance*, *performance* en lijnpatronen in de *valueStreams* en *performance* het kunnen onderscheiden lastig maakt en daardoor ook het beredeneren. In de nieuwe notatie is de *valueStream* een pijl, de *importance* de uitroeptekens en dikte van de pijl wat duidelijk te onderscheiden valt van de *performance* elementen.

Bij de algemene test voor de *problem solving tasks* geeft alleen vraag 2 (Te hoge baktijd) een hogere score voor de nieuwe notatie t.o.v. de oude. Naast het feit dat de *valueStream* en *importance* elementen duidelijk beter scoren in de nieuwe notatie, menen we dat deze elementen bijdragen aan het leerproces om het domein te leren kennen. Volgens Bera et al. (2019) is het begrijpen van een domein a.d.h.v. een model, een leerproces. We zien dat het broodbak model een significant resultaat toont, dit kan te maken hebben met het feit dat de nog relatief jonge en onervaren respondenten eerder een betere voorstelling kunnen maken met een broodbakproces dan het model getoond in vraag 3 (i.e. Onkosten efficiënter afhandelen). Ondanks dat het model getoond in vraag 3 minder associaties oproept bij de respondenten, scoort deze toch beter in de nieuwe notatie dan in de oude notatie.

Paragraaf 4.7 H-Interpretational efficiency comprehension questions en H-Interpretational efficiency problem solving tasks

De algemene tests voor de *comprehension questions* en *problem solving tasks*, geven geen significante resultaten. De afzonderlijke vragen geven geen significante resultaten af, maar er zijn wel een aantal bijzonderheden te vermelden.

In de *comprehension questions* scoren de elementen *value proposition* en *Competence* iets beter in de nieuwe notatie dan in de oude hoewel hun effectgrootte relatief klein genoemd kan worden, resp. $r = .05$ en $r = .09$. Voor de overige elementen scoort de oude notatie beter. Bij de *problem solving tasks* constateren we dat op vraagniveau de oude notatie beter scoort. Dit is een onverwachte uitkomst, men zou verwachten dat zeker de elementen *valueStream* en *importance* een beter resultaat zouden geven.

Uit de test bleek dat het merendeel van de respondenten die de oude notatie hadden, te maken kregen met veel meer kalibratie momenten dan de respondenten die de nieuwe notatie hadden. 56% van alle kalibraties in *comprehension questions* gebeurde bij de oude notatie. Een respondent raakt dan uit zijn concentratie en moet dan weer opnieuw de vraag lezen. Wat we zagen gebeuren in de opnamen is dat de respondent dan nog heel even kort een poging doet, maar al snel een antwoord geeft. Het viel op dat bij de oude notatie 43% van alle kalibraties begon vlak voor of tegen

het einde van vraag 1 terwijl dat bij de nieuwe notatie in slechts 27% procent van de gevallen gebeurde. Het is niet duidelijk waarom de oude notatie meer kalibratie storingen heeft dan de nieuwe. Een verklaring zou kunnen zijn dat respondenten in de oude notatie vaker afdwalen met hun ogen als gevolg van vermoeidheid die ontstaat om het model te begrijpen, waardoor de *eye-tracking* de afdwalende ogen niet meer kan volgen. Volgens Huang (2009) zal het werkgeheugen meer belast wordt als gevolg van het moeilijk begrijpen van een visualisatie. Wanneer dit lang genoeg aanhoudt zal er vermoeidheid optreden en zal de performance omlaag gaan. Aan de andere kant zou het ook zo kunnen zijn dat er meer informatie uit een visualisatie gehaald kan worden als deze duidelijk is waardoor de respondenten langer met de vraag bezig zijn. Dit zou dan weer resulteren in een langere responstijd.

Paragraaf 4.8 H-Perceived ease of use

De algemene tests voor de *perceived ease of use* geeft een significant resultaat af ook wanneer we inzoomen op de vragen gesteld bij dit onderdeel zien we dat bij drie van de vier vragen een significant resultaat wordt afgegeven.

Respondenten die de nieuwe notatie hadden gaven aan dat het model makkelijk is te lezen en daardoor te gebruiken. We nemen aan dat de leesbaarheid (en daardoor ook het gebruiksgemak) vooral te danken is aan het *valueStream* en *importance* elementen die fungeren als een soort richtingaanwijzers waardoor het duidelijk wordt dat het model van onder naar boven moet worden gelezen.

Paragraaf 4.9 AOI run count en AOI fixatie duur – Score

In deze paragraaf wordt gekeken of er een verband bestaat tussen de AOI fixatie *run count*/duur en de scores behaald door de respondenten in de *comprehension questions* en *problem solving tasks*.

Aangezien de AOI fixatie duur voor de *comprehension questions* geen significant resultaat geeft, wordt dit niet in deze paragraaf meegenomen. Zoals in paragraaf 4.9 is besproken, wordt er een positief lineaire verband verwacht, immers hoe hoger de AOI fixatie *run count* en duur hoe beter de visualisatie wordt begrepen, hoe hoger de score zou moeten zijn. Dit verwachte positief lineaire verband zien we alleen terug in de nieuwe notatie bij de *comprehension questions*. Bij de *problem solving tasks* zien we bij de nieuwe notatie in figuur 4.9.2a een licht negatief lineaire verband tussen de AOI *run count* en de score wat veroorzaakt wordt door één respondent die een hoge *run count* heeft maar een lage score. De oude notatie daarentegen toont wel een positief lineair verband. De oude notatie heeft hogere *run counts* wat kan aangeven dat de visualisatie de vraag beter ondersteunt of m.a.w. de *value proposition* wordt door de respondenten eerder opgemerkt als 'niet op zijn plek' zijnde. Wanneer dat zo is moet dit ook zichtbaar zijn in de score van de *comprehension questions* bij de vragen 2 (*Valueproposition*) en 6 (*valueStream-Importance-competence*), maar dat is niet zo, sterker nog er wordt door beide notaties niet gescoord. De figuren 4.9.1a en 4.9.2a geven niet het verwachte resultaat weer vanwege bias die aanwezig is onder de respondenten, m.a.w. we kunnen niet zeggen op basis van een totaal aantal *run count's* en totaal behaalde scores over de vragen of een visualisatie in de nieuwe notatie beter begrepen wordt dan de oude of niet. Respondenten scoren op de ene vraag beter dan de andere wat de oorzaak is van de bias. Daarom geven de figuren 4.9.1b en 4.9.2b een beter inzicht en mogen we concluderen dat de significante resultaten uit paragraaf 4.5 voor de *comprehension questions* en *problem solving tasks* en dat afgezet tegen de scores op vraagniveau, het resultaat laten zien wat Bera et al. (2019) in zijn paper beschrijft nl. dat een hoge AOI *run count* i.c.m. goede antwoorden moet wijzen op een begrijpbare visualisatie.

Paragraaf 4.10 Responstijd – Score

In deze paragraaf wordt de responstijd afgezet tegen de behaalde scores in de *comprehension questions* en *problem solving tasks*. We verwachten bij efficiëntie een hoge score en een korte responstijd, dit zien we wel bij de *problem solving tasks*, maar bij de *comprehension questions* zien we dit niet, althans niet bij de nieuwe notatie. Daar zien we het tegenovergestelde gebeuren, nl. een hoge responstijd tegen een hoge score. Zonder significante resultaten kunnen we geen uitspraken doen over de test. Wel is er een vermoeden dat de iets duidelijkere visualisatie van de nieuwe notatie er voor zorgt dat de respondenten betere aanknopingspunten kunnen vinden dan in de oude notatie. Het gevolg hiervan is dat respondenten langer blijven hangen bij een bepaalde vraag voordat ze antwoord geven. We zien dit verschijnsel ook terug in de veel hogere AOI *run counts* en duur in de nieuwe t.o.v. de oude notatie.

5.2. Reflectie

5.2.1. Interne validiteit

Verstorende factoren voor de interne validiteit zijn respondenten met een oogafwijking als bijv. een amblyopie (lui-oog) of bril, waarbij dit laatste niet zeker is of dit invloed heeft op de kwaliteit van de *eye-tracking*. In hoofdstuk 3 werd de oplossing voorgesteld om alleen respondenten die geen bril droegen uit te nodigen. Na het inrichten van de *eye-tracking* software is met deze software een proef gedaan met een proefpersoon om te zien of er afwijkingen zijn met de *eye-tracking* software. Men heeft deze persoon een visualisatie gegeven die één keer met bril bekeken moest worden en een keer zonder bril. Daarbij moest de persoon iedere keer met de ogen hetzelfde patroon afleggen in de visualisatie. Na het vergelijken van de resultaten bleek dat er geen verschil was tussen het dragen van een bril en zonder bril, wel had de software bij het dragen van een bril meer tijd nodig om te kalibreren.

Een andere verstorende factor wat in hoofdstuk 3 werd aangekaart is dat de respondent zijn hoofd zo stil mogelijk moet houden. Er werd in dat hoofdstuk als oplossing gegeven om kin ondersteuning te gebruiken, maar dit was logistiek niet te regelen. De *eye-tracking* software zelf was 'slim' genoeg om, wanneer een respondent zijn hoofd niet goed hield, opnieuw te kalibreren. De respondent kreeg dan een scherm met twee cirkeltjes waar hij/zij naar moest kijken. De software kon zich dan weer opnieuw instellen waarna de respondent het experiment kon vervolgen. Daarnaast zijn er, voor dat de respondent aan het experiment begon, instructies gegeven over hoe ze moeten omgaan met de *eye-tracking* software (hoofd stil houden, zorg ervoor dat je comfortabel zit, etc.).

Uit de schermopnamen bleek dat het merendeel van de respondenten tijdens het uitvoeren van het experiment tussendoor opnieuw moesten kalibreren wat een verstorende factor is voor de interne validiteit omdat de responstijden van de *questionnaire* software doorlopen tijdens het kalibreren. Dit is opgelost door de kalibratietijd van de responstijd af te halen zodat alleen de effectieve responstijd overblijft.

In hoofdstuk 3 wordt vermoeidheid aangestipt als mogelijke verstorende factor van de interne validiteit. Omdat het hele experiment uit vier losse secties bestaat was het voor de respondent mogelijk tussen de secties door een pauze te nemen of om zelfs de volgende dag verder te gaan. Er zat geen tijdslimiet op de secties behalve bij de *eye-tracking* secties 2 en 3 deze duurde ieder max. 20 minuten.

Een laatste verstorende factor voor de interne validiteit is het feit dat respondenten een zekere mate van dyslexie kunnen vertonen. Het lezen van de vraag duurt dan relatief langer. In hoofdstuk 3 werd aangegeven deze vraag op te nemen in sectie 1 voor de demografische variabelen. Dit is niet gedaan omdat dit opgelost is door de vragen in te korten en het lettertype en grootte aan te passen voor respondenten met dyslexie. Volgens De Bree (2018) is het lettertype Arial en lettergrootte 14 goed te lezen voor mensen met dyslexie. Een proefpersoon met dyslexie heeft het experiment uitgevoerd en kon bevestigen dat het leesgemak van voldoende niveau is.

5.2.2. Betrouwbaarheid

In elk onderzoek komen sterke en zwakke punten naar voren. Dit geldt niet op de minste plaats voor dit onderzoek. Er is in dit onderzoek gebruik gemaakt van een online *eye-tracking service software*. Op een aantal zaken was deze software beperkt. Het eerste zwakke punt hierin was dat het niet mogelijk was om de scroll bewegingen van de respondenten vast te leggen. De respondenten moesten nogal veel scrollen wat er voor zorgde dat een aangegeven AOI wegliep wanneer de respondent ging scrollen. Dit is opgelost door elke opname frame-voor-frame te annoteren, maar hierdoor kan niet worden gegarandeerd dat alle onnauwkeurigheden eruit zijn gefilterd. Het tweede zwakke punt is dat het met de software niet mogelijk was om de anonieme respondenten te koppelen aan de secties 1 (demografische vragen) en 4 (*perceived ease of use*) waardoor het niet mogelijk was om resultaten van bepaalde respondenten af te zetten tegen de resultaten uit de secties 1 en 2. Het derde zwakke punt was het verlies van data door de *eye-tracking* software. Wanneer respondenten klaar waren met het *eye-tracking* gedeelte, werd er niet altijd goed afgesloten, ondanks het feit dat er duidelijke instructies werden gegeven. Het vierde zwakke punt was dat door de Corona *lock-down* de eerste ronde online werd afgenomen waardoor we niet voldoende toezicht konden houden en tussendoor niet de gelegenheid hadden om aanwijzingen te geven. Hiermee had het ontstaan van punt drie voor een groot deel voorkomen kunnen worden. Voor dit onderzoek zaten we net boven het minimum aantal respondenten voor een *Between Subjects* experiment waardoor we de statistische testen met de Mann-Whitney test hebben uitgevoerd omdat deze robuust is voor kleine steekproeven zoals in dit onderzoek is gebruikt. Dit betekent echter wel dat met het tonen van een regressie lijn als in bijv. paragraaf 4.9 en 4.10 kleine afwijkingen gevolgen hebben in zo'n regressie, nl. één respondent kan de richting van zo'n lijn bepalen.

De *questionnaire* werd voordat deze aan de respondenten werd gegeven getest door een PGA deskundige om te controleren of de vragen valide zijn. Er werden twee proefpersonen aangesteld om het experiment te testen, nl. of de ingestelde tijden van max. 20 minuten bij de *comprehension questions* en *problem solving tasks* i.c.m. *eye-tracking* haalbaar is ook voor respondenten met dyslexie en of het experiment met *eye-tracking* voldoende haalbaar is voor brildragende respondenten. Mocht dit nl. niet zo zijn zou de kans groot zijn dat een respondent met dyslexie niet genoeg tijd heeft te antwoorden of de *eye-tracking* data onbetrouwbaar wordt en daarmee het gevaar vormt om verkeerde metingen te krijgen en dus verkeerde conclusies kan trekken.

Aan de respondenten werd een beloning uitgekeerd in de vorm van een gratis lunch bon op voorwaarde dat alle vier de secties (Demografische vragen, *comprehension questions*, *problem solving tasks* en *perceived ease of use*) volledig waren afgerond. De verstorende factor (verstorende factor type: deelnemersvertekening of *bias*) dat hierachter schuilgaat is het feit dat respondenten zomaar wat gaan invullen en niet serieus bezig zijn met het experiment. Door de schermopnamen te bestuderen is gelet op het kijkpatroon dat de respondent aflegt. Bijv. kijkt een respondent alleen naar de vraag en gaat hij/zij vervolgens direct naar de antwoorden zonder naar het model te kijken

of leest de respondent überhaupt niet de vraag en gaat deze direct een antwoord geven. Wanneer dit soort situaties aan de hand waren, werden deze eruit gehaald.

5.3. Conclusies

In dit onderzoek wordt antwoord gegeven op de onderzoeksvraag: Hoe kan de intuïtiviteit van verschillende versies van de PGA notatie statistisch worden vergeleken?

Om het antwoord op de hoofdvraag te geven worden eerst de deelvragen beantwoord:

5.3.1. Deelvraag 1: Wat is intuïtie en waar bestaat deze uit?

Hogarth (2003) beschrijft twee denksystemen, nl. het *deliberate* en het *tacit* systeem. Volgens Hogarth (2003) is met name dit laatste systeem verantwoordelijk van wat wij noemen “intuïtie”. Wij mensen gebruiken onze zintuigen en bouwen door de jaren heen ervaringen op. Het *tacit* systeem zorgt ervoor dat al deze ervaringen en prikkels, samengevat als stimuli, die we binnenkrijgen tijdens ons leven onbewust worden opgeslagen in het langetermijngeheugen. Michael & Mayr (2017) halen nog een andere definitie aan voor intuïtie, nl. *understandability*. Doordat tijdens ons leven informatie bewust en onbewust wordt opgeslagen zorgen de verbindingen die we leggen met ons langetermijngeheugen ervoor dat wij mensen in staat zijn te begrijpen.

5.3.2. Deelvraag 2: Hoe kan intuïtie gemeten en gekwantificeerd worden?

Volgens Hogarth (2003) is de kwaliteit van onze ‘intuïtie’ gebaseerd op interacties/ervaringen die we hebben met de wereld om ons heen. Omdat denkprocessen, dus ook van het tacit denksysteem, zich vertaalt in acties, kunnen we stellen dat we in staat zijn de kwaliteit van intuïtie te operationaliseren door deze acties in een gecontroleerde omgeving te meten. De zgn. *comprehension tasks*, *problem solving tasks* worden vaak ingezet om m.b.v. deze acties de denkprocessen te meten waaronder ook het *tacit* systeem. Het meten van de denkprocessen volstaat niet door alleen gebruik te maken van *questionnaires* waar respondenten hun antwoorden in kunnen geven. We willen ook weten hoe het antwoord tot stand is gekomen, daarom is er naast het vastleggen van de score gebruik gemaakt van *eye-tracking* waarbij we de fixatie en saccades (oogbewegingen) parameters gebruiken om de denkprocessen te meten. Zo kan een hoog aantal saccades duiden op het zoeken naar herkenbare punten in een niet-begrijpbare visualisatie of juist duiden op het vinden van veel relevante gebieden in de visualisatie. Samen met de behaalde scores is het mogelijk de mate van intuïtiviteit te meten.

5.3.3. Deelvraag 3: Hoe kan de relatie gelegd worden tussen intuïtie en de visuele perceptie van conceptuele modelleertalen?

De literatuur beschrijft dat indien een model geen duidelijke structuur heeft of informatie bevat die op meerdere manieren geïnterpreteerd kan worden dit de begrijpelijkheid van een model ondermijnt. Moody (2009) beschrijft in zijn “Physics” of Notations negen principes voor het ontwerpen van effectieve cognitieve visuele notatie. Mendling & Recker (2008) gaan in op het gebruik van tekstlabels in combinatie met pictogrammen en hoe deze samen kunnen leiden tot het

bevorderen van begrijpelijkheid in een conceptueel model. De data afkomstig van de eye-tracking en overige experimenten i.c.m. de gevonden literatuur uit dit onderzoek, zullen een beeld geven over de begrijpelijkheid of intuïtiviteit van de nieuwe t.o.v. oude PGA notatie.

5.3.4. Deelvraag 4: Hoe kunnen de gevonden resultaten uit het theoretisch kader vertaald worden in een concreet uitvoerbare methode?

Respondenten waren eerste- en tweedejaars studenten Bedrijfskundige informatica aan de Avans hogeschool in Den Bosch. De test bestond uit vier onderdelen: (1) Demografische vragen, (2) *Comprehension Questions*, (3) *Problem Solving tasks* en (4) *Perceived ease of use*. Secties 2 en 3 (i.e. *comprehension* en *problem solving tasks*) werden met eye-tracking uitgevoerd. Het experiment is opgezet als een Between Subjects experiment die bestond uit twee groepen, groep A met de oude notatie en B met de nieuwe notatie. De groepen A en B kregen elk dezelfde *comprehension questions* vragen en bij de *problem solving tasks* dezelfde opdrachten. Na de experimenten zijn de resultaten vergeleken en geanalyseerd. De *comprehension questions* bestonden uit meerkeuzevragen waarbij de antwoorden direct te vinden waren in het PGA model. Dit was niet het geval bij de *problem solving tasks*, daar werd een probleem omschreven met een bijbehorend PGA model waarbij de respondenten werden gevraagd de oorzaak van het probleem te vinden en te beredeneren wat de oorzaak zou kunnen zijn van het probleem.

5.3.5. Deelvraag 5: Hoe intuïtief is de nieuw voorgestelde PGA notatie in vergelijking met de oorspronkelijke versie?

Uit de analyse bleek dat de significante resultaten werden gehaald uit de (1) AOI fixatie *run count* van de *comprehension questions*, (2) AOI *run count* en fixatieduur van de *problem solving tasks* en tot slot (3) de *perceived ease of use*. De significante resultaten werden in sterke mate bepaald door slechts een aantal elementen in de nieuwe notatie van de *comprehension questions*, nl. de *valueStream*, *Importance* en *customer goal*. Bij de *problem solving tasks* zijn de opdrachten over de 'Te hoge baktijd' en 'Onkosten efficiënt afhandelen' bepalend voor het significant resultaat. In paragraaf 4.9. werden de AOI *run count* en fixatieduur parameters afgezet tegen de *effectiveness*, uitgedrukt in het aantal scores, bestudeerd. Uit die analyse bleek dat de totale AOI *run count* voor de *comprehension questions* hoger was samen met een hoge score bij de vragen die betrekking hadden op de *valueStream-Importance* en *customer goal*. Bij de *problem solving tasks* zagen we hetzelfde resultaat voor de 'Te hoge baktijd' en 'Onkosten efficiënt afhandelen' problemen. Onze conclusie is dan ook dat de intuïtiviteit in de nieuwe PGA notatie bepaald wordt door de elementen *valueStream*, *Importance* en *customer goal*. Uit de *perceived ease of use* vinden we vooral frustratie bij de respondenten die de oude notatie hadden en was het model in de nieuwe notatie goed te leren lezen en te gebruiken. We schrijven dit dan ook toe aan de hierboven genoemde elementen. Dit zijn ook de elementen die door heel het model gebruikt wordt en dus een verbindende factor heeft in de constructie van een PGA model.

Door de bovenstaande deelvragen te hebben beantwoord kan nu het antwoord worden gegeven op de hoofdvraag,nl:

Hoe kan de intuïtiviteit van verschillende versies van de PGA notatie statistisch worden vergeleken?

Uit het onderzoek van Roelens & Bork (2020) hebben een zestal elementen een verbetervoorstel gekregen wat in dit onderzoek de nieuwe notatie wordt genoemd. In dit onderzoek is gepoogd middels statistische methoden deze oude en nieuwe notatie te evalueren. Hiervoor is naast een *questionnaire* als input bron ook *eye-tracking* software ingezet waarbij deze bronnen zijn gecombineerd. Uit het beantwoorden van de deelvragen is men tot de conclusie gekomen dat statistisch is aangetoond dat de elementen *valueStream*, *Importance* en *customer goal* intuïtief is gebleken in de nieuwe notatie. De *valueStream* en *Importance* zijn vooral ook bepalend voor de begrijpelijkheid van de PGA modelleertaal vanwege hun verbindende rol binnen de taal wat terug te zien is in de resultaten van de *problem solving tasks*. *Customer goal* geniet een zekere mate van intuïtiviteit door haar unieke grafische weergave.

5.4. Aanbevelingen voor de praktijk

Uit dit onderzoek is gebleken dat niet alle elementen van de nieuwe PGA notatie op basis van intuïtiviteit de voorkeur genieten. De elementen *valueStream*, *Importance* en *customer goal* worden als meest intuïtief aangegeven in de nieuwe notatie. Uit paragraaf 5.1.2 zijn een aantal tekortkomingen besproken over de opzet van het experiment. De ervaring leert ons dat het gebruik van een online *eye-tracking* oplossing een aantal nadelen met zich meebrengt. Onze aanbeveling is dan ook om het experiment als volgt op te zetten:

- *eye-tracking* software installeren op de PC of Server in eigen beheer met een aparte camera i.p.v. gebruik te maken van een webcam.
- Zorgen voor een aparte ruimte waar meerdere respondenten tegelijkertijd deel kunnen nemen aan het experiment.
- Op elke plek voor de respondent moet voorzien zijn van een PC met camera waarop de *eye-tracking* software draait.
- Op elke plek zorgen voor verstelbare stoelen zodat de respondent comfortabel kan zitten en evt. een hoofdsteen, waardoor de kans op vermoeidheid klein wordt en daardoor ook de kans op kalibraties.
- Verder is het belangrijk dat de onderzoeker ten tijde van het experiment aanwezig is zodat er vragen kunnen worden gesteld of zodat er ingegrepen kan worden of een houding gecorrigeerd kan worden, of zelfs wanneer er een fout ontstaat in het systeem dit opvangen kan worden door het systeem opnieuw op te laten starten door de onderzoeker.

De secties in dit experiment werden afzonderlijk uitgevoerd en door het gebruik van de online *eye-tracking* software was het niet mogelijk de online *questionnaire* te koppelen met deze *eye-tracking* service. Door gebruik te maken van de *eye-tracking* software geïnstalleerd op een PC of Server in eigen beheer, is het mogelijk de vier secties te integreren in de *eye-tracking* software door de vragen direct op te nemen in deze software.

Verder wordt er geadviseerd de resolutie van de PGA-modellen zo aan te passen dat zij over het gehele scherm te zien is zodat de respondent niet hoeft te scrollen. Hierdoor hoeven er ook geen muisposities worden vastgelegd want deze leveren extra data en complexiteit op wat meegenomen moet worden in de analyse maar wat geen extra toegevoegde waarde geeft. Hier spreekt dan nog

een voordeel uit om de *eye-tracking* in eigen beheer te houden zodat de resolutie afgestemd kan worden op de daarvoor bestemde PC's.

Het verdient tot aanbeveling om bij de elementen van de nieuwe notatie, die niet als intuïtief worden aangemerkt, deze a.d.h.v. de richtlijnen van Moody (2009) zoals beschreven in zijn "*Physics of Notations*" en die van Mendling & Recker (2008) die richtlijnen aangeven voor de tekst bij de iconen, opnieuw te laten ontwerpen door grafische experts. Dit experiment kan dan weer gehanteerd worden om de nieuw ontworpen iconen te evalueren met de oude notatie.

5.5. Aanbevelingen voor verder onderzoek

Het toepassen van *eye-tracking* blijkt een veelbelovend instrument voor het meten van de cognitieve processen tijdens het tonen van een visualisatie van een PGA model. Echter is in dit onderzoek sterk de focus gelegd naar relevante gebieden (*AOIs*). Om een beter begrip te krijgen op wat nu door de respondent als intuïtief wordt ervaren en wat niet, zou er verder onderzoek moeten worden gedaan naar de verschillen tussen de niet-relevante en relevante gebieden. De vraag moet gesteld worden waarom een respondent gebied A relevanter vindt dan het verwachte gebied B. Het verdient de aanbeveling om ook de niet-relevante gebieden te bestuderen in een vervolgonderzoek. In dit onderzoek zijn de vragen/taken gekoppeld aan *AOIs*, men verwacht nl. dat een respondent voor het beantwoorden van een vraag/taak de door de onderzoeker vooraf geannoteerde verwachte *AOI* gebieden binnengaat, dus de relevante gebieden behorende bij de vraag of taak. Wat in dit onderzoek niet wordt meegenomen zijn de gebieden die de respondent nog meer bezoekt buiten de relevante gebieden, de zgn. niet-relevante gebieden. Op het moment dat respondenten bij een vraag of taak consequent bepaalde niet-relevante gebieden bezoeken en er daardoor bijv. een patroon ontstaat verdient dit fenomeen de aandacht om te onderzoeken. Door nu binnen het PGA-model alle mogelijke relevante en niet-relevante gebieden te annoteren, dit kan zowel op icoonniveau als op iconengroep niveau, worden ook de gebieden die niet relevant zijn voor een bepaalde vraag geregistreerd door de *eye-tracking* software. Men zou in de analyse door bijv. *unsupervised machine learning (clustering)* technieken gemeenschappelijke eigenschappen kunnen ontdekken tussen de relevante en niet-relevante gebieden die een relatie kunnen hebben met een bepaalde vraag of vraaggroepen.

Wat niet in dit onderzoek naar voren komt is de mate waarop domeinkennis een rol speelt bij het interpreteren van het PGA-model. Wanneer we bij de *problem solving tasks* de vragen 2 en 3 bekijken resp. 'Te hoge baktijd' en 'Onkosten efficiënt afhandelen', zien we dat het probleem over de baktijd hoger scoort dan vraag 3. Wat al eerder is aangegeven in dit onderzoek, is dat dit te maken kan hebben met de domeinkennis. We kunnen ons voorstellen dat de onervaren respondenten zich beter iets kunnen voorstellen met het baktijd probleem dan onkosten afhandelen. Daarom wordt voor het vervolgonderzoek geadviseerd om het onderzoek te doen binnen een bedrijfscontext waardoor het mogelijk wordt om PGA-modellen te creëren binnen de branche waarin het bedrijf zich bevindt. Respondenten binnen een bedrijfscontext zullen nagenoeg dezelfde domeinkennis bezitten waardoor evt. interpretatie bias kan worden geminimaliseerd.

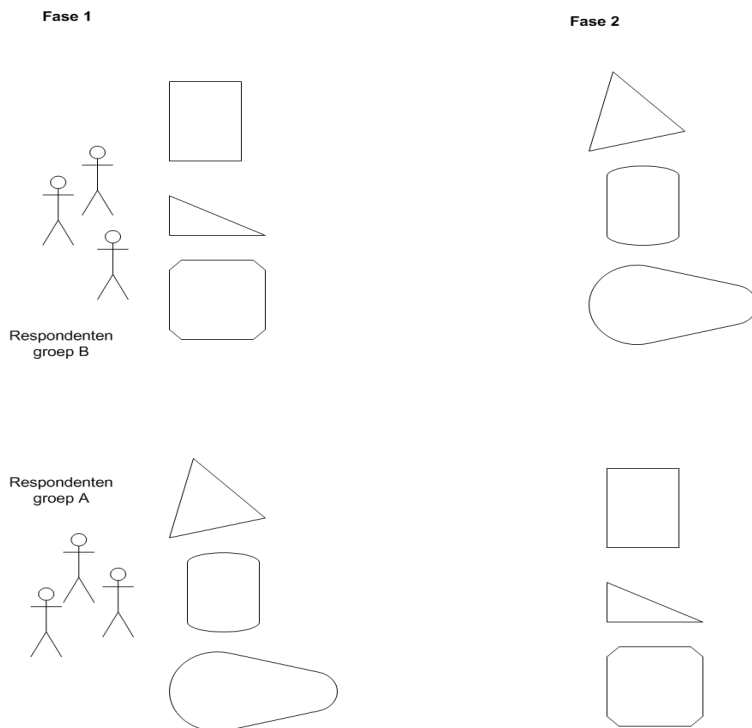
6. Referenties

- Bera, P., Soffer, P., & Parsons, J. (2019). USING EYE TRACKING TO EXPOSE COGNITIVE PROCESSES IN UNDERSTANDING CONCEPTUAL MODELS. *MIS quarterly*, 43(4), 1105. doi:10.25300/MISQ/2019/14163
- Bork, D., Schrüffer, C., & Karagiannis, D. (2019). Intuitive Understanding of Domain-Specific Modeling Languages: Proposition and Application of an Evaluation Technique. In (pp. 311-319). Cham: Springer International Publishing.
- Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yeh, C. S., Borkin, D., . . . Oliva, A. (2016). Beyond Memorability: Visualization Recognition and Recall. *IEEE transactions on visualization and computer graphics*, 22(1), 519-528. doi:10.1109/tvcg.2015.2467732
- Braatz, B., & Brandt, C. (2012). A framework for families of domain-specific modeling languages: Springer-Verlag
- De Bree, E. (2018). Lettertype en leesgemak? Nederlands Kwaliteitinstituut Dyslexie. <https://www.nkd.nl/nieuwsberichten/lettertype-en-leesgemak>
- Hogarth, R. (2003). Educating intuition: a challenge for the 21st century.
- Huang, W., Eades, P., & Hong, S.-H. (2009). Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective. *Information visualization*, 8(3), 139-152. doi:10.1057/ivs.2009.10
- Lankhorst, M. (2013). Enterprise Architecture at Work: Modelling, Communication and Analysis: Springer
- Mendling, J., & Recker, J. (2008). Towards Systematic Usage of Labels and Icons in Business Process Models.
- Michael, J., & Mayr, H. (2017). Intuitive understanding of a modeling language. (Conference Proceedings), 1-10. doi:10.1145/3014812.3014849
- Moody, D. (2009). The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE transactions on software engineering*, 35(6), 756-779. doi:10.1109/TSE.2009.67
- Roelens, B., & Bork, D. (2020). An Evaluation of the Intuitiveness of the PGA Modeling Language Notation. 387, 395-410. doi:10.1007/978-3-030-49418-6_27
- Roelens, B., Steenacker, W., & Poels, G. (2019). Realizing strategic fit within the business architecture: the design of a Process-Goal Alignment modeling and analysis technique: Springer-Verlag
- Saunders, M., Lewis, P., & Thornhill, A. (2016). Research methods for business students.
- The Open Group. (2020). *ArchiMate 3.1 Specification*.

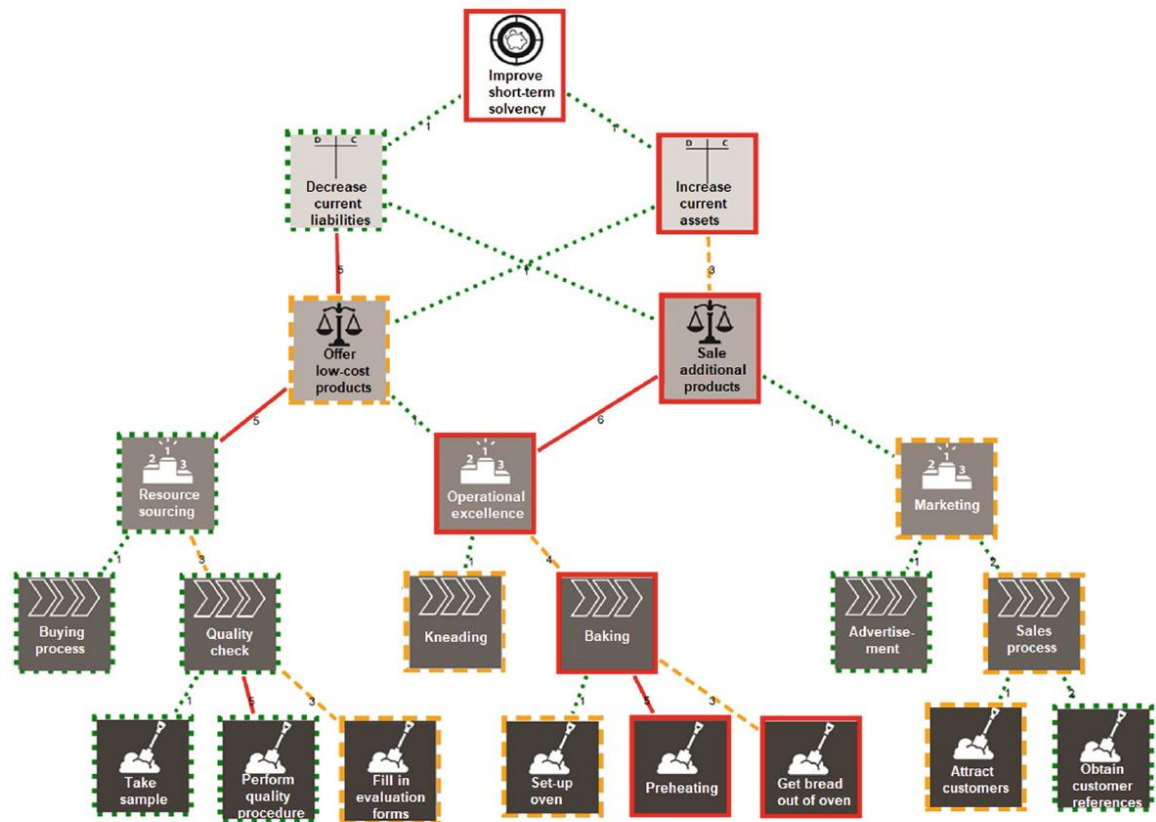
7. Bijlage 1 Opzet experiment van Roelens & Bork (2020)

Opzet experiment voor fase één en twee

Om ervoor te zorgen dat respondenten na bijv. fase 1 de termen die ze hebben gezien niet in fase 2 gebruiken, zijn de elementen en de respondenten in twee groepen verdeeld. In fase één kreeg de eerste groep respondenten de eerste groep van elementen en de tweede groep de eveneens de tweede groep van elementen. In fase twee kreeg de eerste groep van respondenten de tweede groep van elementen en vice versa.



(Fig. 2) Opzet experiment



Legend:

Activity	Process	Competence	Value Proposition	Financial Structure	Internal Goal
Customer Goal	Financial Goal	valueStream	Performance	Importance	
		—			

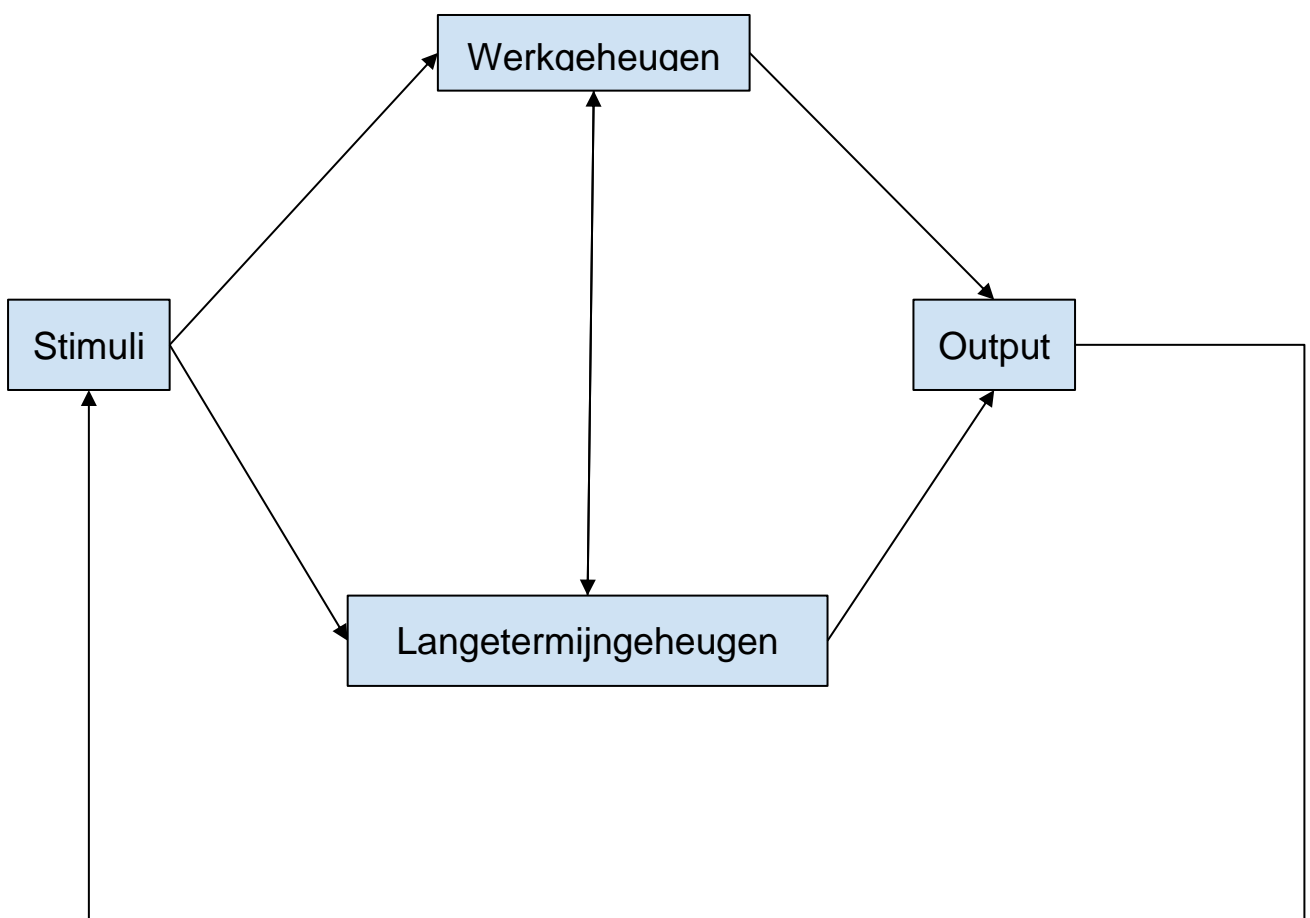
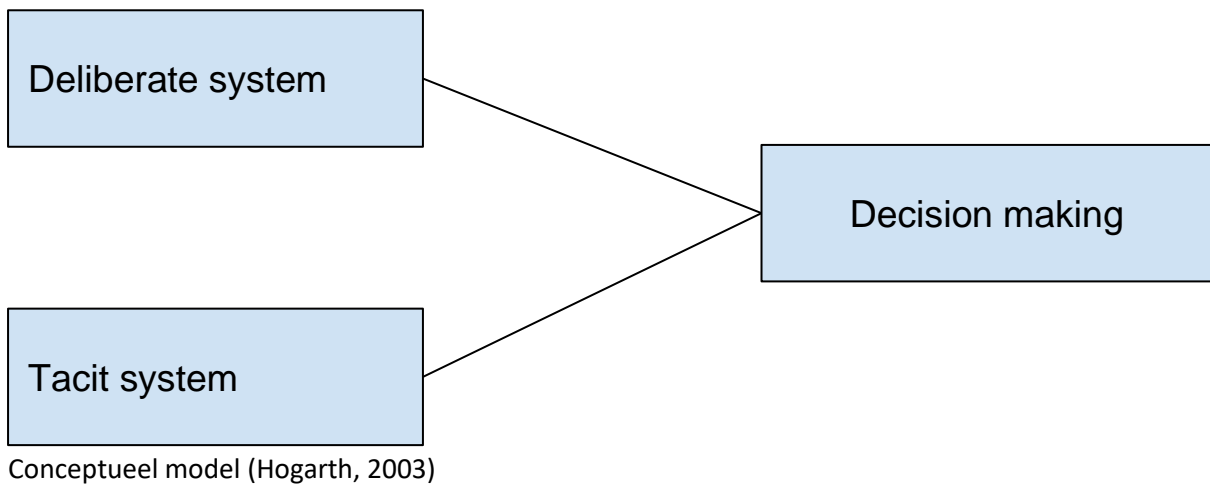
Fig.3 Case Study organisatiestructuur

8. Bijlage 2 Zoek queries voor bronnen

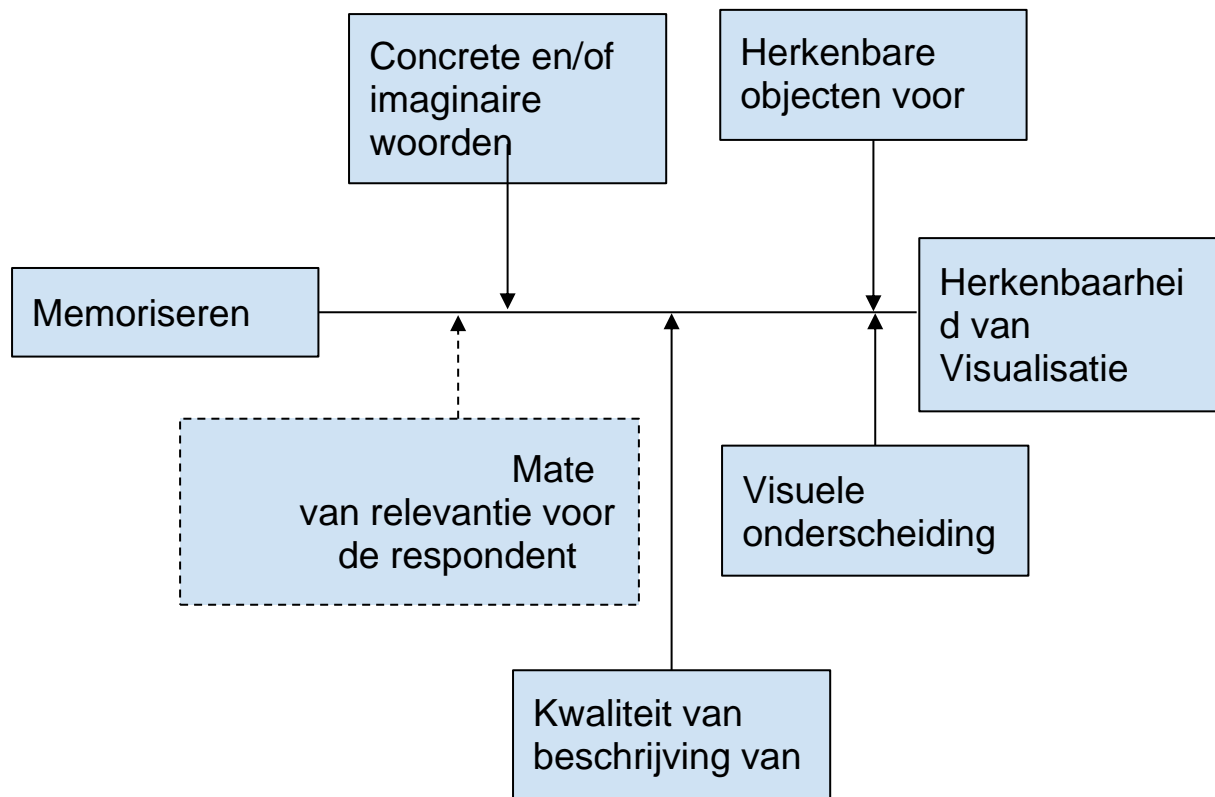
Paper	Paper Title	Publicatie jaar	Ref Web of Science	Link met intuïtiviteit?	Keywords:
1	Intuition, intelligence, data compression	2019	N	J	Intelligence,Intuition,Bounded rationality,Data compression,Complexity theory,Chess,Go
2	Information processing and intuitive decision-making on the fireground: towards a model of expert intuition	2016	J	J	Expertise Intuition Decision-making,Firefighting Cues Information processing
3	Educating Intuition: A Challenge for the 21st Century	2003	J	J	Intuition, Problem Solving
4	Superficial wall stress: the long awaited comprehensive biomechanical parameter to objectify and quantify our intuition	2010	J	J	Intuition
5	Theory-based Bayesian models of inductive learning and reasoning	2006	J Cited by	J	physiology, Concept Formation - physiology, Humans, Problem Solving - physiology, Models, Statistical, Probability Theory, Verbal Learning - physiology, Brain - physiology, Association Learning - physiology, Intuition
6	People Like Logical Truth: Testing the Intuitive Detection of Logical Value in Basic Propositions	2016	J	J	Intuition - physiology, Humans, Japan, Adolescent, Female, Male, Problem Solving - physiology, Decision Making - physiology, Logic, Thinking - physiology, Usage, Intuition, Research, Proposition (Logic), Testing, Index Medicus, Hypothesis testing, Bias, Norms, Conditional probability, Matching, Validity, Hypotheses, Cognition & reasoning, Reasoning, Cognitive style, Bayesian analysis, Heuristic
7	Intuition, reason, and metacognition	2011	J	J	Dual Process Theories, Intuition, Analytic thinking, Reasoning, Retrospective confidence, Metacognition, Psychology, Experimental, Psychology, Social Sciences, Thinking Skills, Psychological Evaluation, Memory, Probability, Data, Models, Robustness (Statistics), Experiments, Statistical Analysis, Intuition - physiology, Humans, Adolescent, Cognition - physiology, Adult, Female, Male, Problem Solving - physiology, Thinking - physiology, Judgment - physiology, Index Medicus
8	Quality Metrics for Information Visualization	2018	N	J	Visual perception, Intuition,quantify
9	ShapeStacks: Learning Vision-Based Physical Intuition for Generalised Object Stacking	2018	N	J	Intuition, visual perception,quantify
10	Learning physical intuition of block towers by example	2018	N	J	Physical intuition
11	Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics	2010	N	N	Visual perception
12	Principles of Contour Information: Reply to Lim and Leek (2012)	2012	J	J	Psychology, Multidisciplinary, Psychology, Social Sciences, New Brunswick, Higher Education, Mathematics Instruction, Probability, Mathematical Concepts, Identification, Foreign Countries, Geometry, Visual Stimuli, Numbers, Theories, Data

					Analysis, Mathematics Education, Rhetoric, Information Theory, Visual Perception - physiology, Form Perception - physiology, Humans, Form perception, Research, Visual perception, Index Medicus, parts, information, shape, contour, curvature
13	Understanding intuition: The case for two forms of intuition	2012	J	J	Inuïtion
14	Intuition: Myth or a Decision-making Tool?	2005	J	J	Inuïtion
15	Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation	1991	J	N	measure perceptions, individual
16	Beyond Memorability: Visualization Recognition and Recall	2016	N	J	Memory, recall, (visual) perception,comprehension,understand
17	Measuring effectiveness of graph visualizations: A cognitive load perspective	2009	J	N	cognitive load, understand, comprehensive
18	The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering (Moody)	2009	J	J	recall,intuition,understandability,comprehensive
19	Using-Eye-Tracking-to-Expose-Cognitive-Processes-in-Understanding-Conceptual-Model	2019	N	J	comprehension, intuitive, problem-solving, task performance, understandability
20	Towards Systematic Usage of Labels and Icons in Business Process Models	2008	J	J	intuition, understandability
21	Eye tracking the Factors of Process Model Comprehension Tasks	2013	N	N	Eye-tracking, relevant regions,comprehension question
	Wat is intuïtie en waar bestaat deze uit?		Query: (measure) AND (intuïtion) AND (Problem Solving)		
	3				
	Hoe kan intuïtie gemeten en gekwantificeerd worden?		1° Query :(intuition) AND (measure) AND (quantify)		2° Query :(visual perception) toegevoegd
	16,17				
	Hoe kan de relatie gelegd worden tussen intuïtie en de visuele perceptie van conceptuele modelleertalen?			Query :(modeling language) AND (eye-tracking)	

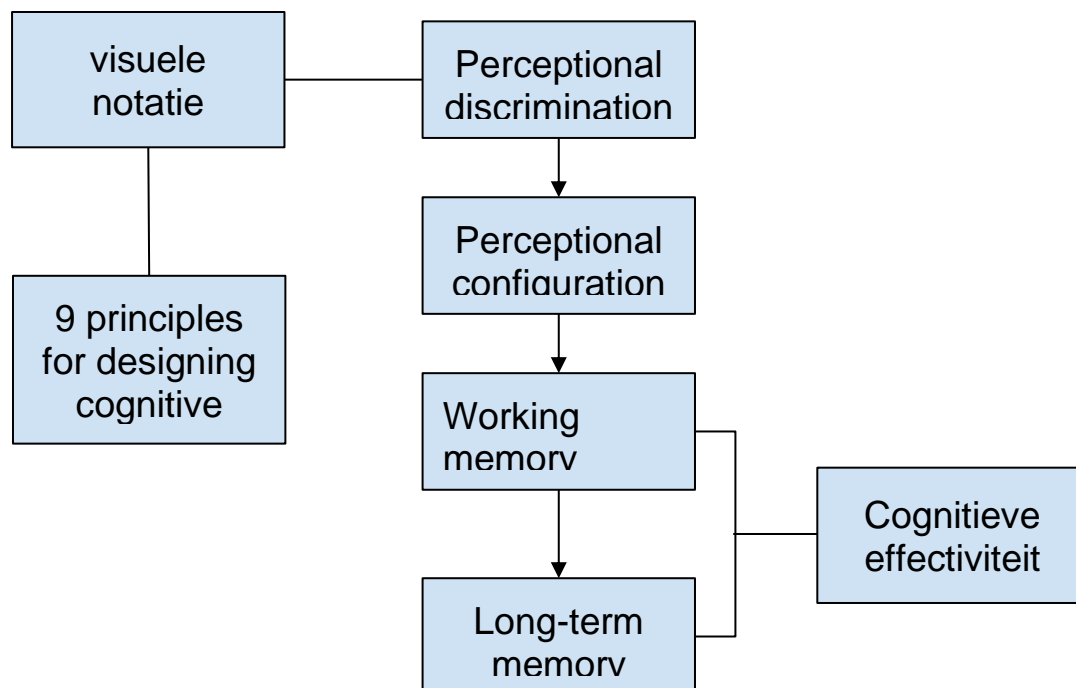
9. Bijlage 3 Conceptuele modellen



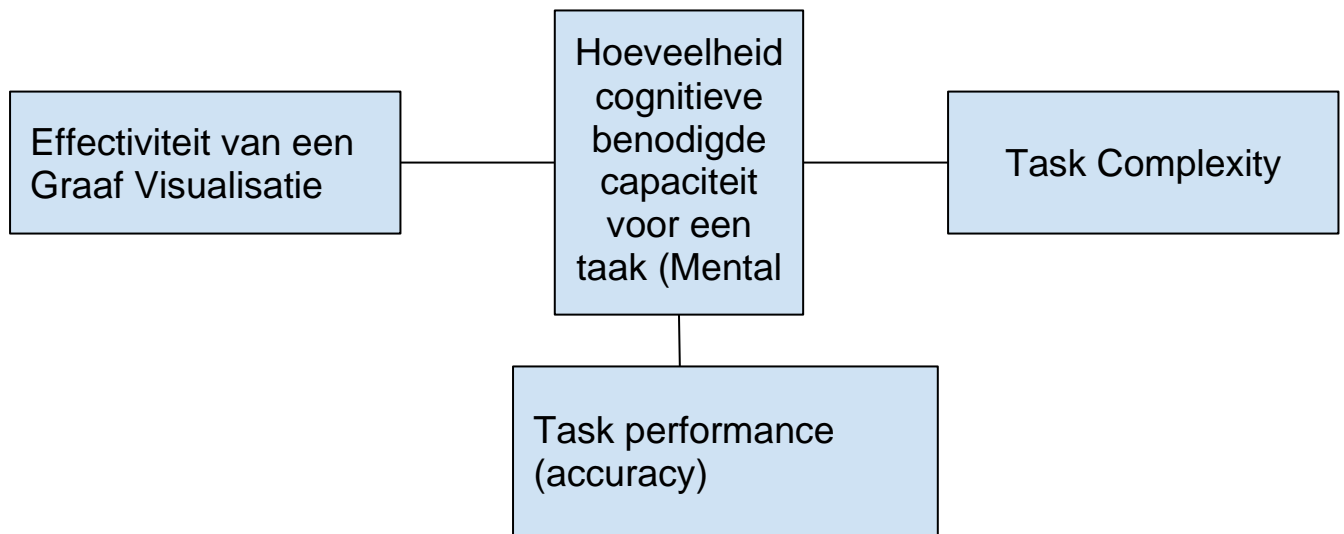
Schematische voorstelling van het denksysteem (Hogarth, 2003)



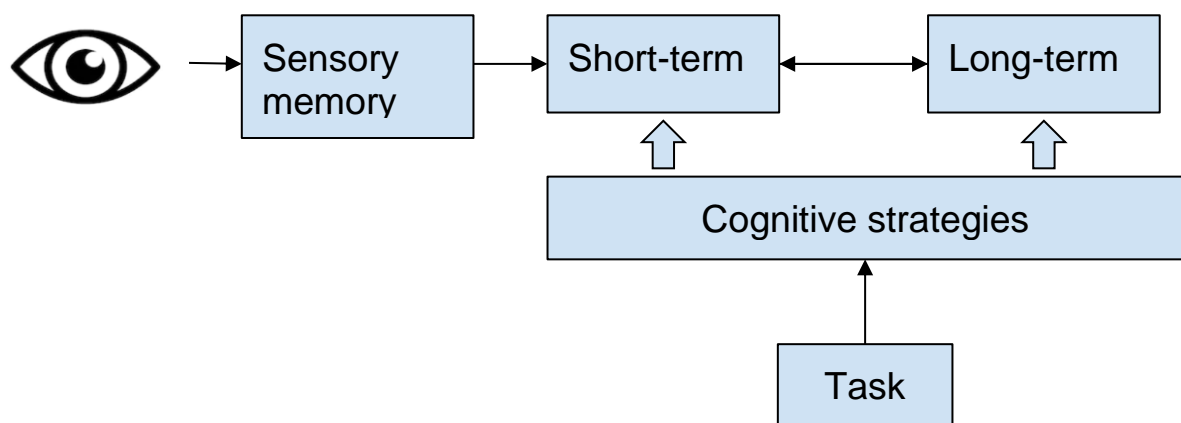
Conceptueel model (Borkin et al., 2016)



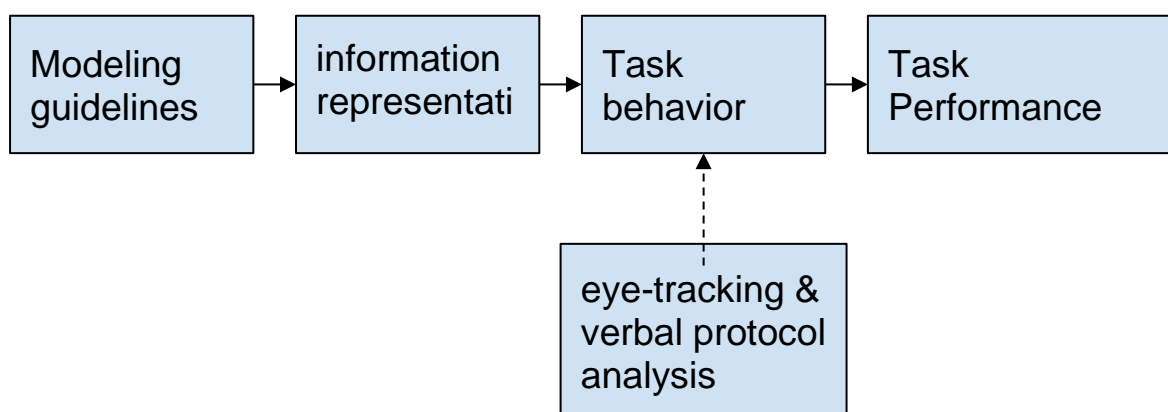
Conceptueel model (Moody, 2009)



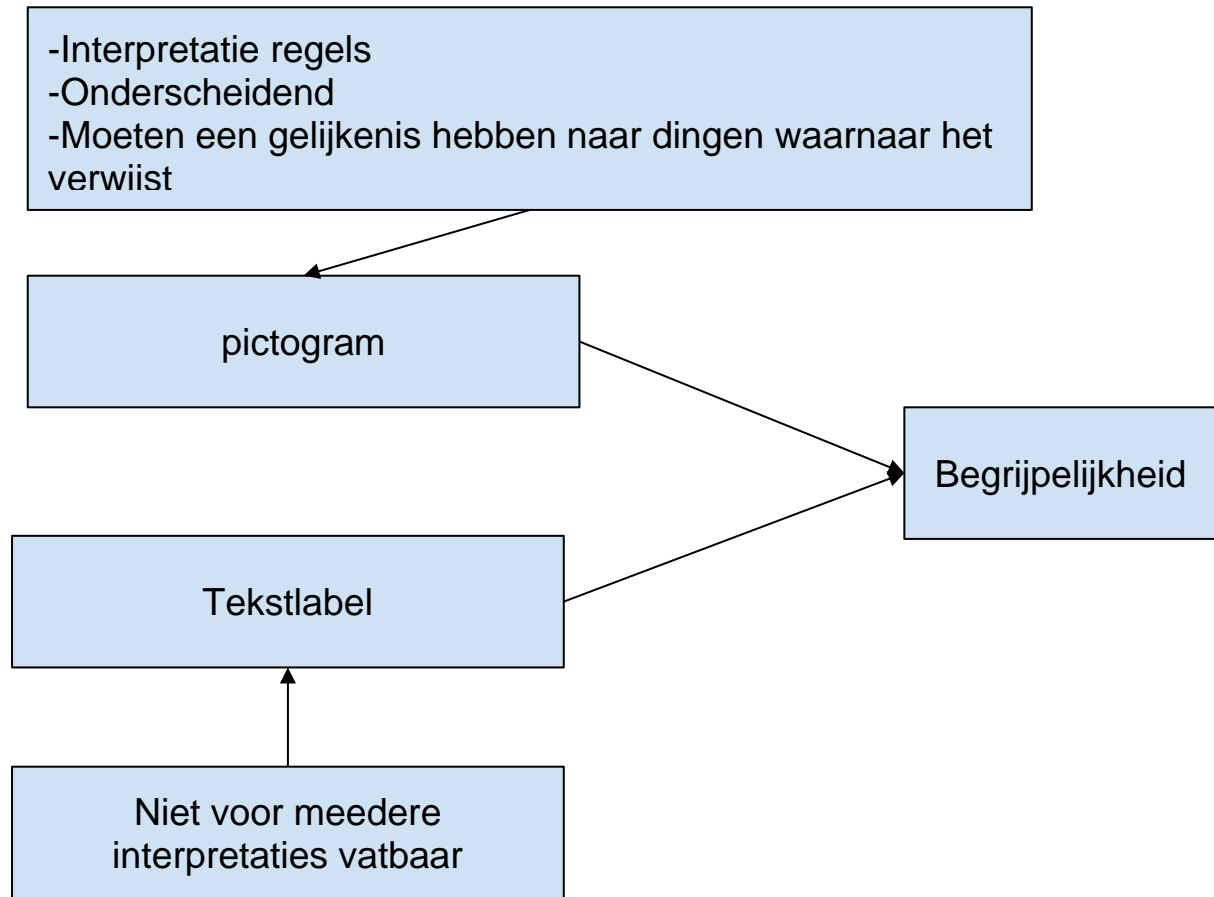
Conceptueel model (Huang et al., 2009)



Schematische voorstelling cognitieve processen (Huang et al., 2009)



In het onderzoeksmodel hieronder is een schematische weergave te vinden (Bera et al., 2019)



(Mendling & Recker, 2008)

10. Bijlage 4 Literatuuranalyse

Analyses:

Beyond Memorability: Visualization Recognition and Recall (Borkin et al., 2016)

Borkin et al. (2016) proberen met deze studie antwoord te vinden op welke elementen in een visualisatie mensen nu eigenlijk letten wanneer ze een visualisatie bestuderen en wat voor informatie nu eigenlijk wordt opgeslagen wanneer ze een visualisatie moeten gaan herkennen.

Hypothesen:

H1: Het memoriseren van een visualisatie heeft een positieve impact op de herkenbaarheid

H2: De kwaliteit van de beschrijving van de visualisatie modereert positief op de herkenbaarheid van een visualisatie

H3: Concrete en/of imaginaire woorden in een visualisatie modereert positief op de herkenbaarheid van een visualisatie

H4: Herkenbare objecten voor de mens modereert positief op de herkenbaarheid van een visualisatie

H5: Visuele onderscheiding modereert positief op de herkenbaarheid van een visualisatie

H6: Mate van relevantie voor de respondent modereert positief op de herkenbaarheid van een visualisatie

Metingen:

Middels *eye-tracking* werd per respondent de data verzameld die uit twee delen bestond. Het eerste deel was het tonen van de visualisatie waarbij de oogfixaties 10 seconden werden vastgelegd. Het tweede deel was het vastleggen van de oogfixaties tijdens het herkennen waarbij de respondenten zelf tekstuele beschrijvingen leverden op basis van de geblurde visualisaties.

Elke respondent kreeg 100 visualisaties te zien willekeurig geselecteerd uit 393 gelabelde visualisaties. Bij het opnemen of bekijken (*encoding phase*) van een visualisatie, werden de x en y coördinaten van de oogbewegingen opgeslagen. De tijd werd gemeten in milliseconden (ms).

Tijdens de herkenningsfase (*recognition phase*) kregen de respondenten de 100 visualisaties wederom getoond plus 100 niet eerder getoonde visualisaties. De 200 visualisaties werden elk 2 seconden willekeurig getoond. Respondenten moesten op de spatiebalk drukken wanneer ze een visualisatie herkende. Ook hier werden de oogfixaties gemeten (ms) samen met de x en y locaties en het aantal correcte antwoorden.

Tijdens de herinneringsfase (*recall phase*) werden de correct beantwoorde visualisaties in een willekeurige volgorde getoond aan de respondent. Dit keer was de visualisatie 60% van het origineel en geblurde zodat evt. tekst niet was te lezen, maar toch voldoende visuele informatie bevatte. Respondenten moesten bij het tonen van een visualisatie in een tekstvak zo nauwkeurig en compleet mogelijk de visualisatie beschrijven van wat ze konden herinneren.

De metingen zijn: oogfixaties (x en y coördinaten), duur van fixaties gemeten in ms, het aantal correcte antwoorden en hoe nauwkeurig een visualisatie door de respondent is beschreven.

Experiment opzet:

Aan het experiment deden 17 mannen en 16 vrouwen mee. De respondenten waren niet kleurenblind en hadden een gemiddelde leeftijd van 22.9 jaar ($SD=4.2$). Elke respondent kreeg een vergoeding van \$25,-. De paper geeft geen informatie over opleidingsniveau. Er wordt wel vermeld dat de respondenten uit de omgeving van Cambridge en Boston komen, daarmee zou het kunnen zijn dat er ook studenten tussen zitten, maar daar kan niet van uit worden gegaan.

Er is sprake van een zgn. within subjects ontwerp omdat alle respondenten dezelfde opdrachten moeten maken. Er is dus geen verdeling tussen de respondenten.

Operationalisatie:

Er zijn 393 gelabelde visualisaties. Daarnaast zijn er ook nog een aantal 'vullers', dit zijn niet eerder getoonde visualisaties die willekeurig worden ingezet tijdens de *recognition phase*. Alle visualisaties werden naar dezelfde dimensies geschaald zodat de maximum dimensie uitkwam op 1000 pixels. Voor de eye-tracking werd de SR Research EyeLink 1000, een desktop eye-tracking systeem met een chin-rest steun gebruikt. De chin-rest steun stond op 22 inches afstand van een CRT monitor (1280 x 1024 pixels). In het begin van het experiment voerden de respondenten een kalibratie uit om het eye-tracking systeem af te stellen. Respondenten konden aangeven wanneer ze even een korte pauze wilde nemen.

Measuring effectiveness of graph visualizations: A cognitive load perspective (Huang et al., 2009)

Om visualisaties en hun performance te meten, dus hoe effectief is een visualisatie eigenlijk, wordt er vaak gegrepen naar metingen als accuraatheid en responstijd. Aangenomen wordt dat een visualisatie het meest effectief is wanneer een respondent in korte tijd weinig fouten heeft gemaakt in het vervullen van een taak of taken. Het komt toch niet vaak voor dat een hoge accuraatheid wordt gehaald in een korte responstijd. Het is meer waarschijnlijk dat een respondent een hoge accuraatheid behaalt, maar tegelijkertijd een lange responstijd heeft door steeds maar weer de afweging te maken tussen een snelle reactie of nauwkeurigheid. Verder wanneer twee verschillende visualisaties worden getoond waarvan de één goed is en de ander niet, is het aannemelijk dat de deelnemer bij de tweede (goede) visualisatie dezelfde cognitieve belasting heeft vanwege de voorgaande (slechte) visualisatie. Performance metingen zullen in dat geval niet het juiste beeld weergeven.

Huang et al. (2009) introduceert twee nieuwe metingen voor het evalueren van (graaf) visualisaties, nl.: *mental effort* en *visualisation efficiency* (Efficiëntie van een (graaf) visualisatie).

Eén van de belangrijke eigenschappen van een effectieve visualisatie is dat het mensen inzicht geeft in informatie die vaak niet direct zichtbaar is. De visualisatie biedt dan met andere woorden mogelijkheden om erover te discussiëren met andere betrokkenen. Het zorgt ervoor dat de visualisatie kan worden begrepen. Om dit te bereiken is het nodig voldoende geheugen capaciteit te hebben wat betekent dat de cognitieve belasting moet worden gereduceerd.

Dit leidt tot de volgende hypothese:

H: *Cognitive Load* en *visualization efficiency* metingen zijn verbeteren op de tekortkomingen van de huidige en gebruikelijke performance metingen.

Experiment:

Men hoopt met het experiment dat wordt opgezet aan te kunnen tonen dat mental effort en visualisation efficiency inzichten te geven die voorheen met de performance metingen niet mogelijk waren.

De deelnemers waren 30 informatica studenten van de universiteit van Sydney. Zij hadden allen ervaring met node-link diagrammen. Elke student kreeg na het afronden van alle taken \$20,-.

Het experiment vond plaats in een computerruimte. Elke student kreeg uit een set van diagrammen een willekeurig gekozen diagram waar hij/zij drie taken mee moest doen.

De volgorde van de taken waren willekeurig gekozen voor elke student. Elke keer wanneer een diagram werd getoond moest de deelnemer zo snel mogelijk een antwoord geven op de vraag die werd getoond. De respons tijd werd gelogd vanaf het moment dat een diagram werd getoond totdat de deelnemer op de stop knop drukte. De deelnemer vult op een 9 punt Likert Scale de ervaren mental effort in, drukt vervolgens weer op de knop om het volgende diagram te laten tonen, waarna de responstijd weer loopt etc..

Metingen zijn *mental effort*, dit is de hoeveelheid cognitieve capaciteit dat is gealloceerd en nodig is om een bepaalde taak uit te voeren. *Visual efficiency* kan worden gedefinieerd als de cognitieve winst t.o.v. de cognitieve kosten, hiermee zien we mental effort plus *response time* als cognitieve kost en cognitieve winst als *accuracy*. *Response time* en *accuracy* zijn waarden afkomstig uit metingen uit bijv. *problem-solving tasks* waarbij het aantal correcte antwoorden worden vastgelegd (*accuracy*) en tijd die er voor nodig was om te komen tot het antwoord (*response time*).

The “Physics” of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering (Moody, 2009)

Moody (2009) beweert in zijn paper dat de visuele vorm bepalend is voor de begrijpbaarheid van het model, vooral door de meer onervaren eindgebruikers. Kleine veranderingen in een model kunnen al nadelig zijn voor de begrijpbaarheid ervan. Moody (2009) introduceert negen principes voor het ontwerpen van effectieve cognitieve visuele notatie.

Semantic Transparency

Eén van deze principes raakt duidelijk de intuïtiviteit van een notatie, nl. *Semantic Transparency*. Hiermee wordt de zgn. *naturalness of intuitiveness* bedoeld van de notatie.

Volgens Moody (2009) wordt door *Semantic Transparency* de cognitieve belasting gereduceerd vanwege de zgn. ingebouwde geheugensteun. Het verdient de voorkeur om functioneel vergelijkbare objecten te gebruiken. Bijv. een vuilnisbak icoontje kan een verwijderen actie betekenen.

Semiotic clarity

Het visuele vocabulaire (grafische symbolen) en de regels hoe deze symbolen te gebruiken (de grammatica) vormen samen de syntax. De grafische symbolen zelf symboliseren de semantische constructies of metamodel. Volgens het *Semiotic clarity* principe moet er een 1-op-1 mapping zijn tussen grafisch symbolen en de semantische constructies.

Perceptual discriminability

Is het gemak waarmee we grafische symbolen kunnen onderscheiden van elkaar.

Moody (2009) geeft aan dat hoe beter symbolen van elkaar zijn te onderscheiden hoe meer wij in staat zijn een visualisatie correct te interpreteren.

Complexity management

Is de mate waarop een visuele notatie zijn corresponderende informatie kan weergeven zonder dat het brein te veel wordt belast.

Principle of Cognitive Integration

Is volgens Moody (2009) alleen maar toepasbaar wanneer er meerdere diagrammen in het spel zijn. Bij cognitieve integratie wordt de informatie van verschillende diagrammen geïntegreerd in het brein om te komen tot een coherent mentaal, dus in het brein opgebouwd, systeem.

Visual expressiveness

Hier wordt onder verstaan het aantal visuele variabelen die in een notatie worden gebruikt.

Bij gebruik van een reeks van visuele variabelen kan dit voor de zgn. *computational offloading* zorgen. Er zijn twee subcategorieën, nl. *Information-carrying variables* variabelen gebruikt om informatie te coderen in een notatie en *free variables* dat zijn niet gebruikte variabelen.

Wanneer we enkel *free variables* hebben, zijn er geen visuele variabelen en spreken we van louter tekstuele informatie. Wanneer we geen *free variables* hebben, spreken we van een verzadigde visualisatie, dus een visualisatie zonder tekst. De meeste SE notaties gebruiken slechts één visuele variabele nl. *shape*.

Dual Coding

Volgens de *dual coding theory* is het gebruik van tekst en grafische informatie samen erg effectief wanneer het gaat over het overbrengen van informatie.

Graphic Economy

Dit aspect is gerelateerd aan het aantal grafische symbolen dat zich in een visualisatie bevindt.

Hoe meer symbolen of hoe groter het vocabulair des te meer wordt de lezer van de visualisatie cognitief belast. Hij/zij moet immers meer symbolen onthouden. Dit kan weer worden opgelost met het aanleggen van een legenda, maar dan moet deze dan ook weer regelmatig worden geraadpleegd wat ook weer een extra belasting levert.

Cognitive Fit

Volgens Moody (2009) zegt de *cognitive fit* theorie dat verschillende visuele representaties ook horen bij verschillende taken en verschillend publiek.

Visuele notaties hebben tot doel ondersteuning te bieden bij communicatie en het oplossen van problemen. Daarbij is het van belang dat cognitieve belasting bij mensen laag blijft. Dus een visualisatie moet cognitief effectief zijn. De hierboven genoemde principes zijn richtlijnen om cognitieve effectiviteit te bereiken. Sommige principes kunnen elkaar tegenspreken bijv. Semiotic Clarity kan Graphic Economy tegenspreken. *Semiotic Clarity* veroorzaakt immers door 1-op-1 mappings met semantische constructies groei in het aantal visuele symbolen van de visualisatie. Hier moet men rekening mee houden bij het ontwerpen van een model.

H: Wanneer een visuele notatie wordt ontworpen volgens de negen principes, wordt de cognitieve effectiviteit vergroot waardoor menselijke communicatie en problem-solving aspecten worden geoptimaliseerd.

Educating Intuition: A Challenge for the 21st Century (Hogarth, 2003)

Volgens Hogarth (2003) zijn er twee denksystemen (*two modes of thought*) Het *tacit* en het *deliberate* systeem. Volgens Hogarth is het *tacit* systeem (wat ook 'intuïtie' wordt genoemd) verantwoordelijk voor het onbewust verwerken van informatie. Dit systeem opereert automatisch en is verantwoordelijk voor het aanmaken van het langetermijngeheugen. Het legt alle interactie en ervaringen vast die wij als mensen hebben als wij contact hebben met de wereld om ons heen. Zie fig. 1 bijlage 3.

Het *deliberate* systeem is het bewuste en analytisch systeem van ons denken. Het veroorzaakt cognitieve belasting vanwege het aanspreken van ons beperkte capaciteit van het werkgeheugen of kortetermijngeheugen.

Volgens Hogarth (2003) werken de twee denksysteem samen tijdens het nemen van bijvoorbeeld beslissingen. Het menselijk denkproces wordt geactiveerd door stimuli, dit kan een visuele of een voel stimulus zijn of bijv. een gedachte die weer een andere gedachte triggert, etc..

Wanneer stimuli het denksysteem binnenkomt wordt onbewust de informatie opgeslagen in het langetermijngeheugen voor evt. toekomstig gebruik. Stimuli kunnen ook direct resulteren in een actie bijv. wanneer je in een drukke winkelstraat loopt waarbij je een constante stroom van stimuli binnen krijgt en toch maar een beetje hoeft op te letten. Daarbij is het interessant te ontdekken dat de actie komt vóór het begrijpen of de bewustwording van de actie.

Het *deliberate* systeem speelt een belangrijke rol wanneer we ons moeten concentreren op een stimulus, bijv. bij het lezen van een boek. Het overruled dan het *tacit* systeem. Mensen zijn in staat om tijdens bewuste processen in het *deliberate* systeem te besluiten of automatische processen de controle over moeten nemen. Bij het besturen van een auto maken we bewuste keuzes waar we heen willen en hoe hard we rijden, maar laten het schakelen bijv. weer over aan automatische processen.

In de meeste gevallen is het *tacit* systeem ons default systeem waarbij het *deliberate* systeem wordt aangeroepen wanneer de default het niet kan oplossen, bijvoorbeeld wanneer er een bewust proces

als het maken van een planning moet plaatsvinden. Volgens Hogarth (2003) werken de twee systeem dan ook bij het maken van beslissingen nauw samen.

Volgens Hogarth (2003) is de kwaliteit van onze intuïtie gebaseerd op de interacties/ervaringen die we hebben met de wereld om ons heen. Het meten van die kwaliteit wordt dan gedaan op basis van het aantal goede beslissingen die gemaakt zijn of de behaalde winst die geboekt wordt door het nemen van bepaalde beslissingen.

Hogarth (2003) beweert dat wanneer mensen hun intuïtie blijven onderwijzen dit uiteindelijk de kwaliteit in het nemen van beslissingen omhoog brengt. Bij het leren speelt het *tacit* systeem een belangrijke rol. Dat wil zeggen dat we door het *tacit* systeem automatisch leren, maar aangezien we niet altijd bewust zijn van dit leren, kan er ook niet altijd gezegd worden of de ervaringen die we ondergaan een goede 'leermeester' zijn. De output van ons denken resulteert in een actie of acties. Hierop ontvangen we feedback wat weer dient als input en als stimulus ons denksysteem binnenkomt. Wanneer de kwaliteit van feedback laag is zal ook de kwaliteit van de intuïtie laag zijn binnen een bepaald domein.

H1: De kwaliteit van intuïtie wordt bepaald door de kwaliteit van feedback die wordt ontvangen en wanneer feedback wordt gegeven.

H2: Onbewuste processen en bewuste processen werken nauw samen bij het nemen van beslissingen.

Using Eye Tracking to Expose Cognitive Processes in Understanding Conceptual Models (Bera et al., 2019)

De paper wil onderzoeken hoe de prestaties van probleemoplossende taken kunnen worden voorspeld wanneer identieke informatie in verschillende diagram presentaties wordt getoond.

Om dit te onderzoeken zijn er twee cognitieve operaties van belang om informatie te kunnen verwerken. *Attention* of aandacht op verschillende delen van het model en het integreren van informatie over verschillende delen van het model die relevant zijn voor het uitvoeren van een taak.

Bij het lezen maakt het oog snelle bewegingen waarbij de aandacht verschuift van het ene deel van de visualisatie naar het andere deel en dan weer stilstaat omdat het brein bezig is het materiaal op die locatie te interpreteren. Dit stilstaan van het oog noemen we een fixatie wat wordt gebruikt om de aandacht te meten die wordt gegeven aan een bepaald deel of element in een visualisatie. Een fixatie duurt tussen de 200 tot 300 milliseconden (ms) en geeft aan dat een lezer aan een bepaald deel op de visualisatie de aandacht vestigt. Naast de duur van een fixatie kennen we ook de fixatie aantal dat het belang aangeeft van een deel of element in een visualisatie.

Met deze parameters worden m.b.v. *eye-tracking* twee soorten aandacht-gerelateerde analyses gedaan: het eerste is de algemene verkenning van de visualisatie en het tweede is de specifieke aandacht gericht op zgn. sub-gebieden of *Areas Of Interest* (AOIs). De parameters geven aan of een

visualisatie alleen gescand wordt of dat er daadwerkelijk informatie tot zich wordt genomen door de lezer.

Met het meten van aandacht alleen is het niet mogelijk iets te kunnen zeggen over de mate van taakprestatie. Volgens Bera et al. (2019) moet de tijd of duur die gespendeerd wordt aan een bepaald gebied in relatie worden gebracht met de taakprestatie.

Cognitive fit en load

De *Cognitive fit* theorie. Deze theorie stelt dat wanneer de informatie die in een weergave wordt benadrukt overeenkomt met het type informatie die nodig is voor een taak, de respondenten hun taak beter uitvoeren dan wanneer deze informatie niet overeenkomt.

De *cognitive load* theorie gaat over de beperkte capaciteit van het werkgeheugen wat een belangrijke rol speelt tijdens het uitvoeren van een probleemoplossende taak. Bera et al. (2019) onderscheiden drie typen van *cognitive load*. (1) *Intrinsic load*, dit wordt bepaald door de complexiteit van de taak (2) *extraneous load*, die wordt bepaald door de informatie die wordt getoond en (3) *germane load*, die wordt veroorzaakt door de behoefte om informatie te integreren i.c.m. kennis afkomstig van het langetermijngeheugen om een taak uit te voeren. Omdat deze drie typen allemaal gebruik maken van de beperkte capaciteit van het werkgeheugen, moet geprobeerd worden de *extraneous load* tot een minimum te beperken en zo ruimte te geven aan de overige twee typen. Door de *cognitive fit* theorie is het volgens Bera et al. (2019) zo dat wanneer taak relevante informatie visueel opvalt in een model, dan zal de lezer zijn aandacht daarop richten waardoor dit een positieve invloed zal hebben op de taakprestatie. Verder zegt de paper dat bij beslissings taken de beslissingsnemers alleen de aandacht vestigen op gebieden in het model met een hoge taak-relevantie en irrelevante taken negeren waardoor de *extraneous load* wordt gereduceerd.

Metten van aandachtsgebieden en associatie van informatie

De relevante aandachtsgebieden kunnen gemeten worden d.m.v. de hierboven beschreven fixatie parameters, maar om de associatie van informatie te meten zijn de fixatie parameters niet voldoende. Bij een conceptueel model beperkt een taak zich niet tot één aandachtsgebied, maar zijn er meerdere gebieden die bij elkaar horen die zorgen voor de complete informatie die nodig is om de taak goed uit te voeren (associatie van informatie). Bij meerdere aandachtsgebieden zoals deze voor kunnen komen in een model, zal het oog regelmatig afwisselen tussen fixaties, genaamd *saccades*. Uit verschillende bronnen die de paper aanhaalt wordt aangegeven dat het meten van *saccades* wordt gebruikt om de mate van associatie van informatie weer te geven. De paper stelt echter dat hiervoor nooit een empirisch bewijs is geleverd, maar wil dat in deze paper wel gaan vaststellen middels *eye-tracking*. De parameters om de *saccades* te meten zijn (1) *run count*, geeft het aantal keren aan dat een bepaalde AOI a.h.w. is binnengekomen en weer is verlaten (2) *AOI run count*, is bijna hetzelfde als de *run count* met dat verschil dat hier de afwisselingen wordt gemeten tussen twee of meer AOIs.

Experimenten opzet algemeen:

Experimenten 1 en 2 hebben een between-subjects ontwerp waarbij elke respondent een aantal modellen kreeg toegewezen. De respondenten moesten m.b.v. het model probleemoplossende vragen beantwoorden. Een hoge score van goede antwoorden gaf aan dat veel informatie was geïntegreerd in het langere-termijn geheugen en dus een diep niveau van begrip aanwezig is.

Experiment 1:

De respondenten waren 45 masterstudenten Informatiesystemen van een Universiteit uit zuiden van de VS, zij waren verdeeld in drie groepen van 15 studenten, hierna te noemen respondenten.

Elke respondent werd ingelicht over het *eye-tracking* systeem. Elk experiment bevat drie stadia. Het eerste stadium was een kalibratie sessie voor het *eye-tracking* systeem waarbij op het scherm een reeks puntjes werden getoont. In het tweede stadium werd de respondent een aantal vragen gesteld over zijn/haar ervaring met domein modelleren. Het derde stadium bestond uit het uitvoeren van twee probleemoplossende taken voor twee modellen gepresenteerd in verschillende talen: (1) *Event-driven Process Chains* (EPC) en (2) *Business Process Model Notation* (BPMN). Tijdens het uitvoeren van de taken volgde het *eye-tracking* systeem hun oogbewegingen.

Experiment 1 heeft vier rol gerelateerde taken waarbij de verwachting is dat BPMN het beste scoort. Dit heeft alles te maken met het feit dat BPMN rollen beter kan structureren. Met *lanes* is het bijv. mogelijk een grove indeling te maken en daar weer in verder op te splitsen mocht dat nodig zijn. BPMN geeft de mogelijkheid om op basis van taak relevantie te structureren dit in tegenstelling tot EPC waarin alles als een ketting aan elkaar is geregen. EPC is hierdoor ook meer geschikt voor sequentie gerelateerde taken dan bijv. rol gebaseerde taken. In dit experiment wordt een aanpassing gemaakt in het EPC model. Delen die rol gerelateerde taken bevatten worden gemarkeerd (visueel opvallend). Elke rol wordt gemarkeerd met een andere kleur. In het experiment wordt dit model EPC *Highlighted* genoemd (EPC-H).

In het experiment worden 4 hypothesen getest:

H1: Vergeleken met de gebruikers van EPC, zullen gebruikers van EPC-H modellen meer gericht zijn op taak relevante gebieden om een rol gerelateerde probleemoplossende taak uit te voeren

H2: Vergeleken met de gebruikers van EPC-H, zullen gebruikers van BPMN modellen meer gericht zijn op taak relevante gebieden om een rol gerelateerde probleemoplossende taak uit te voeren.

H3: Prestaties in een rolgerelateerde probleemoplossende taak kunnen worden verklaard door de mate waarin aandacht is voor taakrelevante gebieden.

H4: Prestaties in een rolgerelateerde probleemoplossende taak kunnen worden verklaard door de visuele associatie van informatie tussen taakrelevante gebieden.

Experiment 2

Dit experiment werd uitgevoerd door 27 *business* bachelorstudenten uit een universiteit uit het midwesten van de Verenigde Staten. Deze studenten hadden allemaal kennis van bedrijfsprocessen door een cursus SAP Enterprise Resource Planning te volgen waar het onderwerp werd geïntroduceerd. Zij kregen training in de concepten van bedrijfsprocessen drie uur voordat het experiment begon.

De focus van dit experiment lag in het herhalen van experiment 1 met gebruik van EPC en EPC-H. Het is van belang te benadrukken dat BPMN hier niet wordt gebruikt. Het BPMN model bevatte maar één centraal gebied waar de rollen zich bevonden in de visualisatie. Hierdoor hoefde de ogen niet af te wisselen over meerdere gebieden in het model en dus waren de saccadische parameters niet relevant. Het waren vooral de fixatie waarden die hier uitsprongen.

Hoewel in experiment 1 is aangetoond dat er een link is met taakprestatie en visuele associatie kan er nog niet gesproken worden over cognitieve integratie. Cognitieve integratie is de synthese van informatie uit één of meerdere bronnen. In dit experiment zijn deze bronnen verschillende delen binnen een model of concepten binnen en buiten het model (voorkennis). In dit experiment willen men cognitieve integratie testen met *eye-tracking* en *verbal protocol analysis* (VPA). VPA is een methode om cognitieve processen te analyseren tijdens het interpreteren van modellen en kan dus gebruikt worden om te begrijpen hoe verschillende modellen invloed hebben op de taakprestaties. Aan de respondenten werd gevraagd om tijdens het uitvoeren van hun taken hardop te denken zodat er inzicht verkregen kan worden in hun denkprocessen.

Om deze cognitieve integratie te operationaliseren worden de verbale uitspraken tijdens het uitvoeren van taken geanalyseerd. De VPAs van de respondenten werden omgezet naar tekst en overhandigd aan twee codeerders, die overigens niets afwisten van het doel van het experiment. De codeerders hadden een lijst waarop termen stonden die in het model voorkwamen. Een codeer document over cognitieve integratie werd aan de codeerders overhandigd. Hierop stond de taak beschreven en de mogelijke antwoorden per taak. De VPAs werden door de codeerders afzonderlijk gecodeerd en kwamen samen om de verschillen te bespreken.

Towards Systematic Usage of Labels and Icons in Business Process Models

Mendling & Recker (2008)

De paper bespreekt een systematische aanpak in het gebruik van pictogrammen en labels van in de activiteiten in een procesmodel en hoe deze kunnen bijdragen tot verbetering in het begrijpen van procesmodellen.

De paper baseert zich op de *Dual Coding Theory*, die beweert dat wanneer individuen informatie verwerken, zij gebruik maken van twee kanalen, nl. het visuele en het auditieve kanaal (i.e. bij het lezen van tekst speelt het auditieve gedeelte in de hersenen een rol). Deze kanalen vullen elkaar aan zodat de ontvangen informatie bijdraagt aan het begrijpen van deze informatie. Wanneer informatie alleen door één kanaal wordt ontvangen zal de mate van begrip vele malen minder zijn.

Om de twee kanalen (visueel en auditief) in te kunnen zetten om zo te komen tot het beter kunnen begrijpen van een procesmodel, moet een pictogram een mapping of beter gezegd een directe relatie hebben naar zijn tekstlabel. Het probleem is echter dat sommige van deze tekstlabels op meerdere manieren geïnterpreteerd kunnen worden waardoor het juist moeilijk wordt een model te begrijpen. De paper stelt daarom een aanpak voor om dit probleem op te lossen door een zgn. werkwoord classificatie schema (WCS) te introduceren binnen de context van het procesdomein. In deze paper zijn er twee WCSs schema's bestudeerd, nl. het *MIT Process Handbook* en de *Levin verb Classes*.

Als voorbeeld hebben de onderzoekers deze WCSs toegepast op de labels van de activiteiten in het SAP referentie model, dat meer dan 3000 procesmodellen bevat over de SAP R/3 functionaliteit ter ondersteuning van bedrijfsprocessen in organisaties.

Verb Occurrences		Verb Occurrences		Verb Occurrences	
process	2003	post	330	update	203
enter	1922	release	328	analyze	191
determine	1755	maintain	316	settle	186
check	971	calculate	271	allocate	180
create	665	assign	261	transmit	171
plan	614	define	258	copy	164
transfer	510	edit	258	print	162
select	349	perform	228	generate	141
confirm	345	specify	226	change	136
carry out	337	evaluate	203	display	131

Figuur 1 SAP referentie model

In fig. 1 staan de meest voorkomende 30 werkwoorden afgebeeld uit het SAP referentie model.

- ⊕ [Create](#)
- ⊕ [Modify](#)
- ⊕ [Preserve](#)
- ⊕ [Destroy](#)
- ⊕ [Combine](#)
- ⊕ [Separate](#)
- ⊕ [Decide](#)
- ⊕ [Manage](#)

Figuur 2 Generieke werkwoorden *MIT Process Handbook*

In fig. 2 staan acht generieke werkwoorden afkomstig van de *MIT Process Handbook*. Wat de onderzoekers hebben gedaan is, m.b.v. *WordNet*, alle werkwoorden van *MIT Process Handbook* gelinkt met het referentie model van SAP en in kaart gebracht hoeveel synoniemen er gevonden zijn per generiek werkwoord. M.a.w., hoe vaak komen de generieke werkwoorden voor in de referentie model van SAP? Anders dan bij het *MIT Process Handbook*, zijn de werkwoord klassen afgeleid uit linguïstische analyses en gelinkt met het reference model van SAP. Elk werkwoord klasse is weer

onderverdeeld in sub-classes waarbij iedere klasse een lijst van meest voorkomende werkwoord voorbeelden heeft.

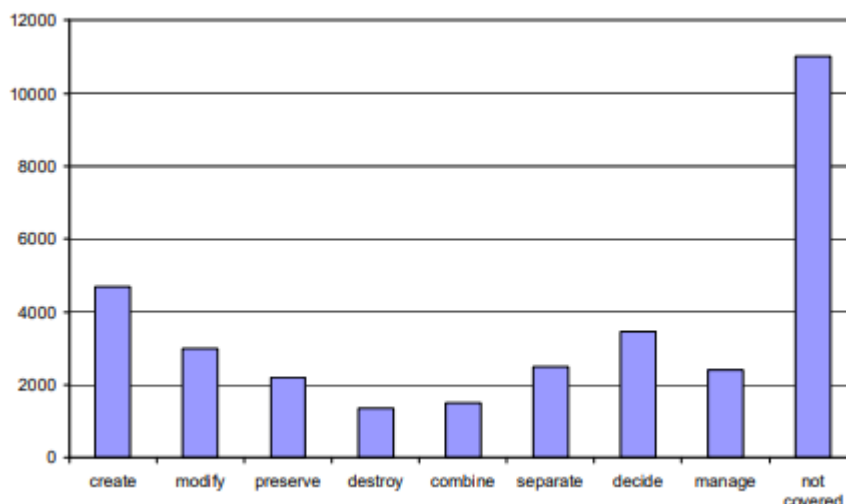
01.4.2 "with" Preposition Drop Alternation

agree argue banter bargain battle bicker box brawl chat
chatter chitchat clash coexist collaborate collide combat
commiserate communicate compete concur confabulate confer
conflict consort consult converse cooperate correspond court
cuddle date debate dicker differ disagree dispute dissent
divorce duel elope embrace feud fight flirt gab gossip haggle
hobnob hug jest joke joust kiss marry mate meet mingle mix
neck negotiate nuzzle pair pass pet play plot quarrel quibble
rap rendezvous schmooze scuffle skirmish spar spat speak
spoon squabble struggle talk tilt tussle vie visit war
wrangle wrestle yak

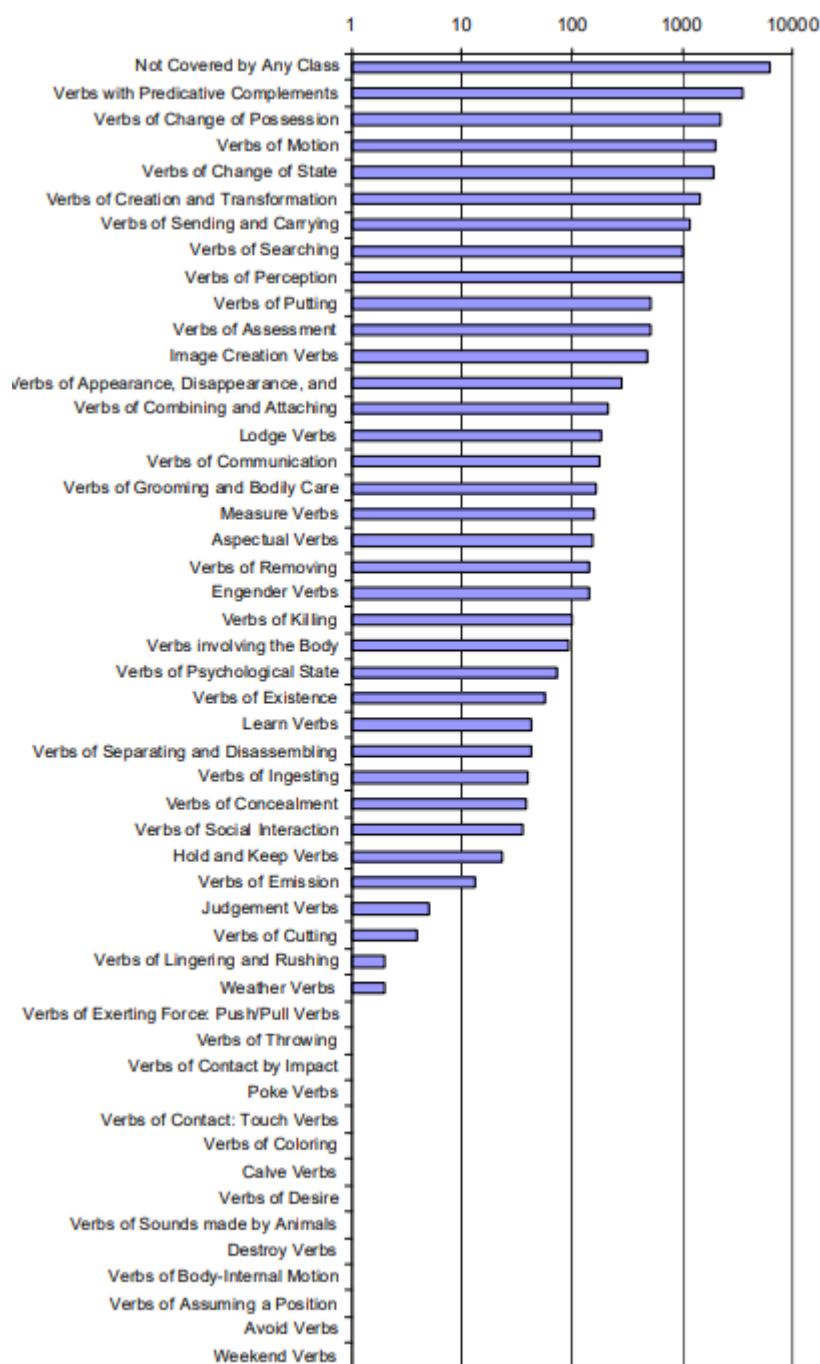
02.1 Dative Alternation

accord acknowledge address adjudge adjudicate administer
admit adopt advance allege allocate allot announce anoint
appoint articulate ask assert assign assume avow award babble
baptize barge bark bash bat bawl bear begrudge believe bellow
bequeath bet bill bleat bode boom bounce brand bray bring
broadcast bunt burble bus cable cackle call carol carry cart
catapult cede chant charge chatter chirp christen chuck cite
cluck communicate concede confess consecrate consider
contribute convey coo cost credit croak croon crow crown cry
declare decree deem delegate deliver demonstrate denounce
deny describe designate dictate dispatch display distribute
donate drag drawl drive drone drop dub e-mail elect elucidate
entrust envy esteem exhibit explain explicate express extend
fancy fax feed ferry find fine flash flick fling flip float
fly forbid forfeit forgive forward furnish gabble gibber give

Figuur 3 de *Levin verb classes*



Figuur 4 Het aantal voorkomende synoniemen per generiek werkwoord van de SAP referentiemodel gelinkt met het *MIT Process Handbook*.



Figuur 5 Het aantal voorkomende synoniemen per generiek werkwoord van de SAP referentiemodel gelinkt met *Levin verb classes*

Het resultaat was dat de *Levin verb Classes* 68% en het *MIT Process Handbook* 44% van het SAP referentie model bedekte. Hierdoor viel deze laatste af en zijn de onderzoekers doorggegaan met het onderzoek op basis van de *Levin verb Classes*.

De werkwoorden werden vervolgens aan pictogrammen gekoppeld. Deze werkwijze voorkomt volgens de paper dat pictogrammen worden gekoppeld aan tekstlabels die op meerdere manieren geïnterpreteerd kunnen worden. Een pictogram dat op zichzelf intuïtief goed gekozen is kan,

wanneer deze wordt gekoppeld aan een label dat op meerdere manieren geïnterpreteerd kan worden, de begrijpbaarheid van het model alles behalve verbeteren.

Het ontwerpen van pictogrammen is volgens de onderzoekers meer een kunstvorm dan wetenschap. Toch zijn er richtlijnen waaraan een pictogram moet voldoen:

- 1) pictogrammen moeten een gelijkenis hebben naar dingen of taken waarnaar het verwijst
- 2) ze moeten grafisch duidelijk te onderscheiden zijn
- 3) compositie principe: regels voor het samenstellen van pictogrammen moeten natuurlijk en gemakkelijk te begrijpen en te leren zijn. De *Multiple-level icon composition principle* heeft regels voor het samenstellen van hoe een pictogram op hoog niveau moet worden samengesteld uit de lagere niveau pictogrammen.
- 4) interpretatieregels: regels voor het samenstellen van pictogrammen moeten overdraagbaar zijn op verschillende modellen en doelgroepen.

Dit leidt uiteindelijk tot de volgende hypothese:

H:Wanneer pictogrammen voldoen aan ontwerprichtlijnen en tekstlabels zijn niet voor meerdere interpretaties vatbaar, verhoogt dit de begrijpelijkheid van een conceptueel model

11. Bijlage 5 Data en preparatie

11.1. Algemeen (weghalen *outliers*)

Het onderzoek is opgedeeld in vier secties die elk een afzonderlijke dataset vormen te weten: Sectie 1 Demografische vragen, Sectie 2 *Comprehension Questions*, Sectie 3 *Problem Solving tasks* en Sectie 4 *Perceived ease of use*.

In *Comprehension Questions* en *Problem Solving tasks* wordt met eye-tracking gewerkt. Hieronder worden de stappen beschreven die zijn ondernomen om te komen tot bruikbare data voor analyse.

De eye-tracking data bestaat uit een aantal onderdelen, nl. gegevens over de duur van de fixatie, de start- en eindtijd van de fixatie en de positie (x, y coördinaten) van de fixatie (zie figuur 5.1a) en AOI gebieden die handmatig zijn gelabeld met de fixatie duur en count (figuur 5.1b).

tester_id	fixation_point_x	fixation_point_y	fixation_starts_at_ms	fixation_ends_at_ms	fixation_duration_ms
c0cb4a3e-2787-457d-a53e-296194c33a9	47,268518518519	23,854166666667	47373	47487	114
c0cb4a3e-2787-457d-a53e-296194c33a9	39,114583333333	33,645833333333	47798	47988	190
c0cb4a3e-2787-457d-a53e-296194c33a9	42,708333333333	36,875	48472	48713	241
c0cb4a3e-2787-457d-a53e-296194c33a9	60,729166666667	52,864583333333	49151	49369	218

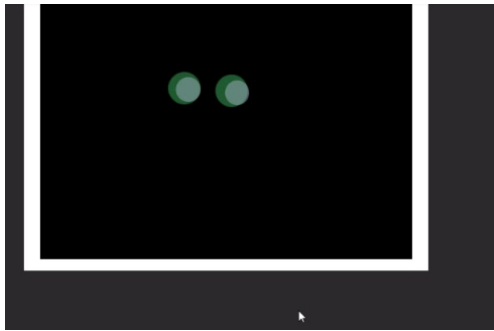
Figuur 5.1a

aoi_name	aoi_fixation_total_count	aoi_fixation_average_duration_ms
Fvraag1	2	151
Fvraag1	2	117
Fvraag1	1	104
Fvraag2	1	230
Fvraag2	1	138
Fvraag2	2	184
Gvraag1	1	124
Gvraag1	1	206
VerbindingA-Bvraag2	1	238
VerbindingB-Avraag1	1	166
VerbindingB-Avraag1	1	216
VerbindingB-Avraag2	2	189
VerbindingC-Bvraag1	2	213
VerbindingC-Bvraag1	2	160
VerbindingC-Bvraag1	2	166
VerbindingC-Bvraag1	2	203
VerbindingC-Bvraag1	1	122
VerbindingC-Bvraag1	1	304
VerbindingC-Bvraag2	2	163
VerbindingE-Dvraag1	1	213

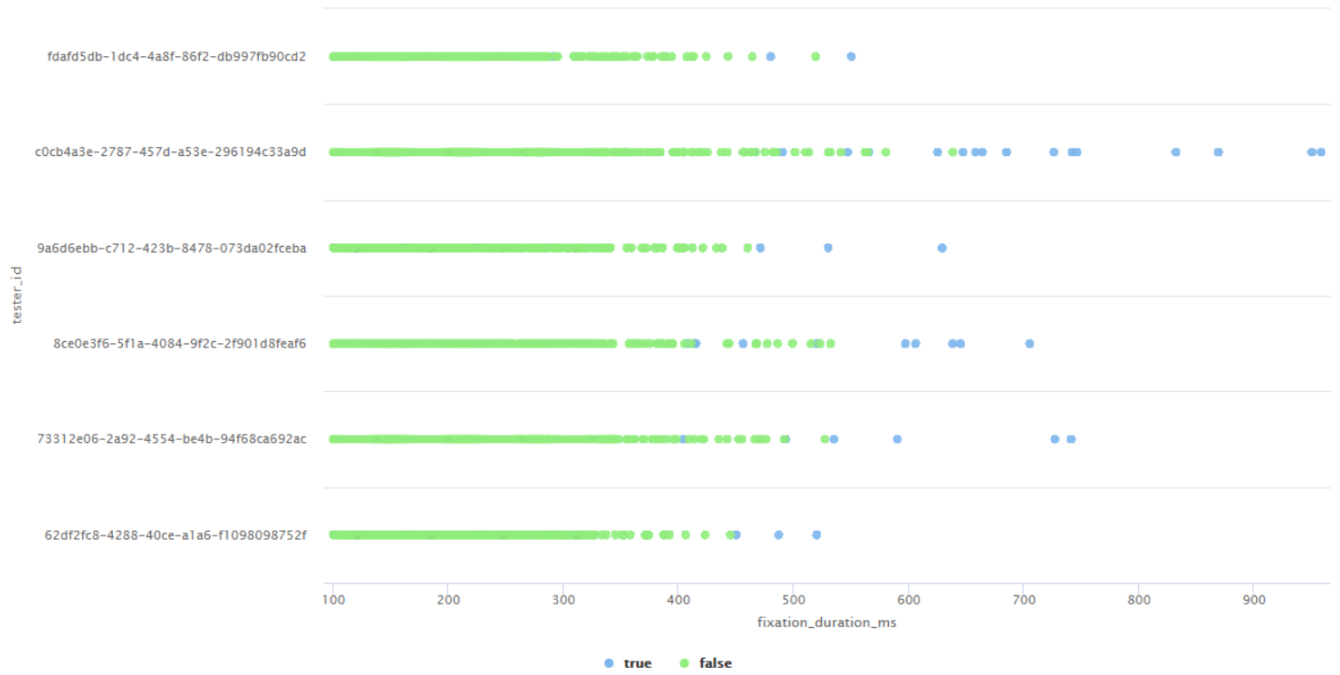
Figuur 5.1b

Allereerst zijn de zgn. *outliers* eruit gehaald. In eerste instantie zijn er handmatig twee respondenten uit het onderzoek gehaald. Dit is gedaan door alle opnamen één-voor-één te bekijken. Hierdoor werd er ontdekt dat één respondent alleen de antwoorden bekeek en zomaar wat invulde en nummer twee had extreme x coördinaten. Uit de opname bleek dat deze respondent twee schermen open had staan met in één scherm het onderzoek en het andere scherm een Netflix film. Het was duidelijk te zien dat de ogen meer dan de helft van de tijd gericht waren op het scherm waar de film op stond dan het andere scherm met het onderzoek. Na deze filtering zijn de overige *outliers* eruit gehaald uit beide groepen deze bestonden uit twee stappen. Tijdens het onderzoek kwam het regelmatig voor dat de *eye-tracking* software soms de ogen van een respondent niet kon volgen. Wat er dan gebeurd is dat de software opnieuw moet gaan kalibreren. Dit gebeurd dan middels een scherm (zie figuur 5.2) waarbij de respondent zijn ogen moet richten op twee cirkeltjes, wanneer de ogen daar precies opvallen kan het systeem weer verder gaan. Echter de seconden lopen wel door, daarom is eerst gekozen om alle kalibratie momenten van alle respondenten uit de data te halen. Dit is ook gedaan met de responstijden van Lymesurvey, nl. wanneer een respondent een dergelijk kalibratie scherm voor zich krijgt loopt de tijd in Lymesurvey door waardoor er verschillen ontstaan.

De laatste stap is het verwijderen van overige uitbijters (*outliers*). Deze werden gekenmerkt door extreme waarden. Deze waarden halen het gemiddelde uit balans en geeft daardoor een minder goed beeld van de data en zijn verdeling. Figuren 5.3, 5.4, 5.5 en 5.6 laten de *outliers* zien van de nieuwe notatie en figuren 5.7, 5.8, 5.9 en 5.10 de *outliers* van de oude notatie.



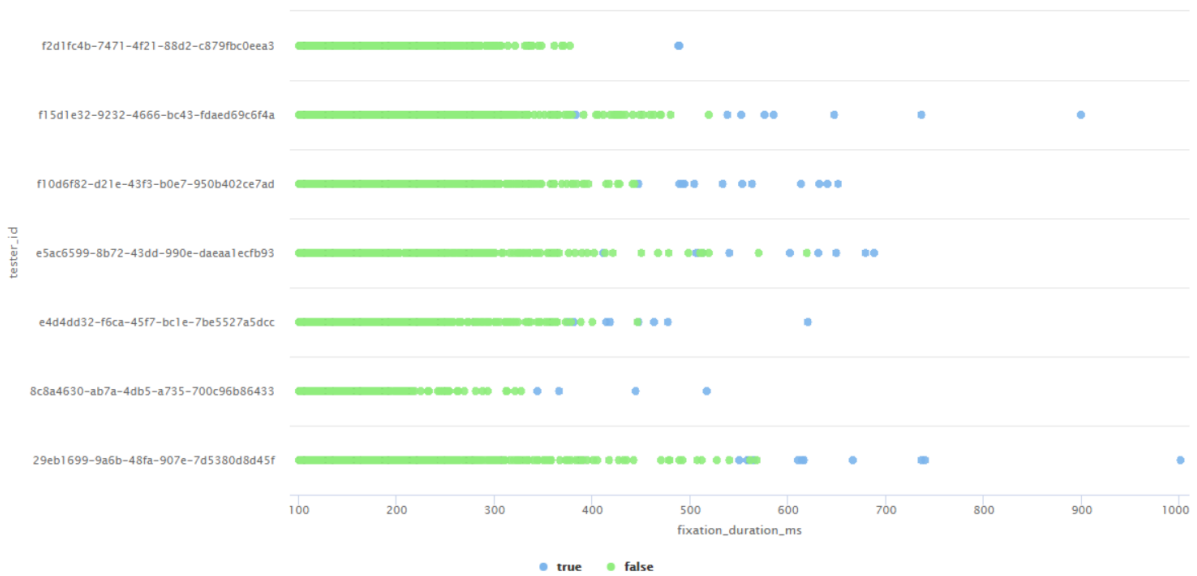
Figuur 5.2 Kalibratie scherm



Figuur 5.3 De blauwe punten zijn de *outliers* sectie 2.



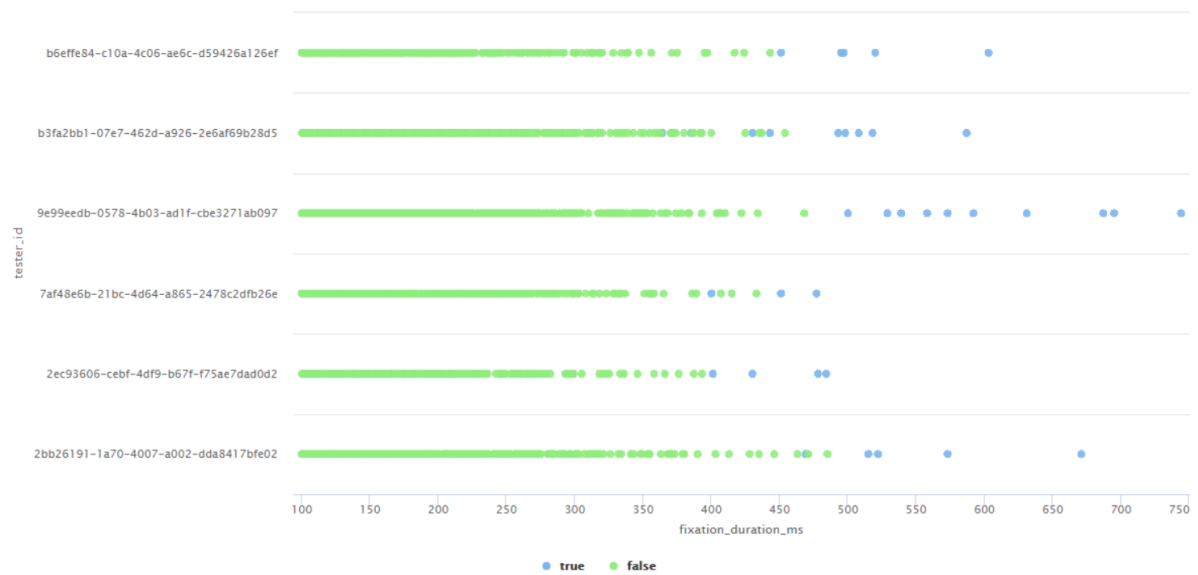
Figuur 5.4 De blauwe punten zijn de *outliers* sectie 2.



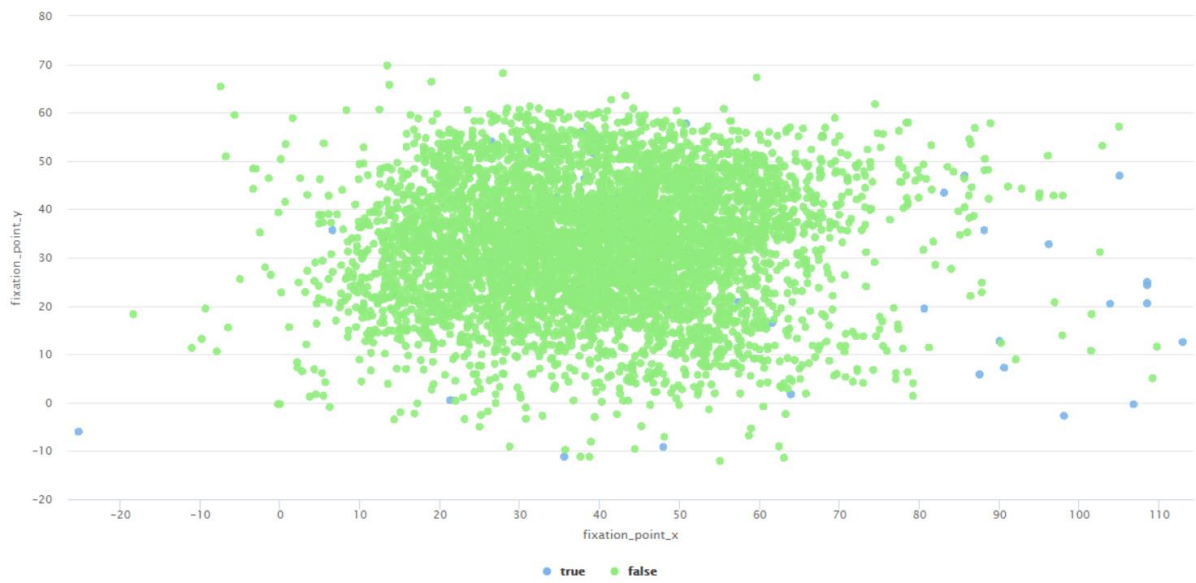
Figuur 5.5 De blauwe punten zijn de *outliers* sectie 3.



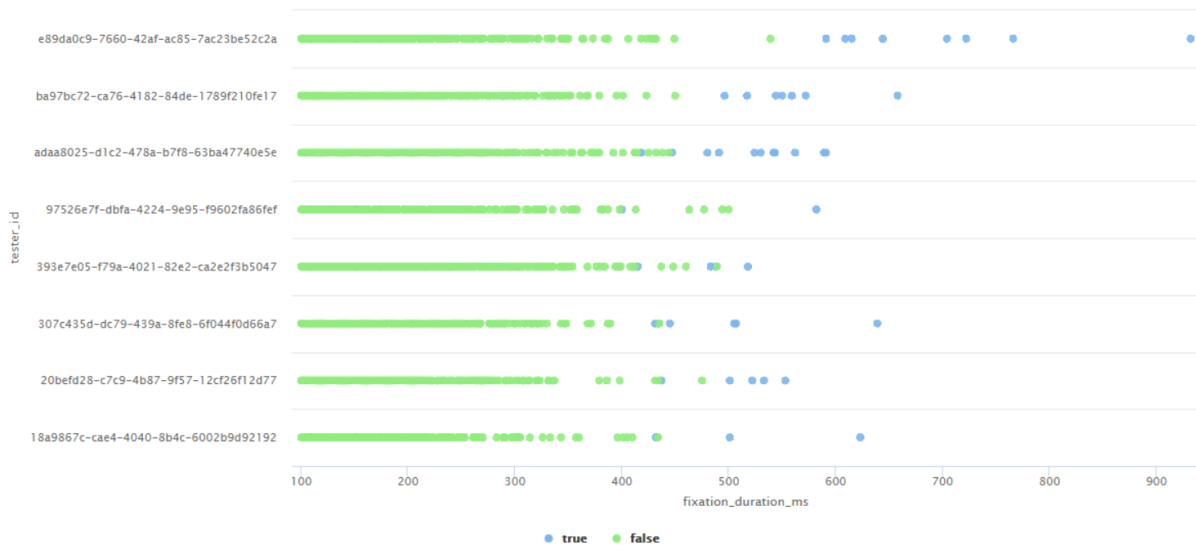
Figuur 5.6 De blauwe punten zijn de *outliers* sectie 3.



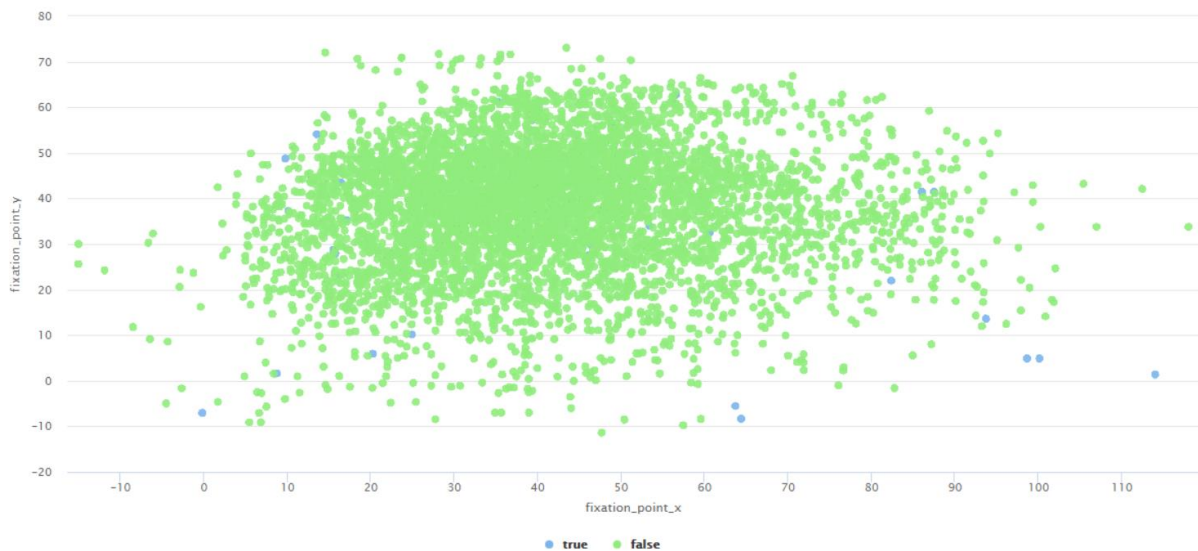
Figuur 5.7 De blauwe punten zijn de *outliers* sectie 2.



Figuur 5.8 De blauwe punten zijn de *outliers* sectie 2.



Figuur 5.9 De blauwe punten zijn de *outliers* sectie 3.



Figuur 5.10 De blauwe punten zijn de *outliers* sectie 3.

11.2. Sectie *comprehension questions* en *problem solving tasks* data preparatie algehele fixatie duur en aantal

De data is zodanig aangepast dat deze bruikbaar is om antwoord te kunnen geven op de hypothesen H_a -fixatie aantal en H_a -fixatie duur.

fixation_duration_ms	AantalVanFixation_duration_ms	SomVanfixation_duration_ms	Notationvers
100	30	3000	0
101	31	3131	0
102	41	4182	0
103	55	5665	0
104	37	3848	0
105	50	5250	0
106	42	4452	0
107	42	4494	0
108	56	6048	0
109	66	7194	0
110	49	5390	0
111	53	5883	0
112	48	5376	0
113	31	3503	0
114	59	6726	0
115	20	2200	1

Figuur 5.11

In figuur 5.11 is een frequentie tabel te zien. *Fixation_duration_ms* zijn de afzonderlijke fixation_durations, fixatie intervallen, die voorkomen waarbij het tweede veld *AantalVanFixation_duration_ms* de frequentie aangeeft hoe vaak een bepaalde fixatie duur voorkomt. *SomVanfixation_duration_ms* geeft aan hoeveel er totaal is gespendeerd aan fixaties van bijv. 100 ms in de steekproef. M.a.w.

$SomVanfixation_duration_ms = Fixation_duration_ms * AantalVanFixation_duration_ms$

Het veld *Notationversion* is de categorische variabele met de waarde '0' voor oude notatie en '1' voor de nieuwe notatie.

Controle op normale verdeling

Er is voor deze dataset de Shapiro-Wilk test uitgevoerd om na te gaan of de variabelen een normale distributie volgen. Figuren 5.12 en 5.13 tonen de resultaten van de test voor respectievelijk sectie 2 en 3

Tests of Normality							
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Notationversion	Statistic	df	Sig.	Statistic	df	Sig.
AantalVanfixation_duration_ms	0	,138	6	,200*	,985	6	,974
	1	,172	6	,200*	,932	6	,594
SomVanfixation_duration_ms	0	,167	6	,200*	,982	6	,963
	1	,155	6	,200*	,949	6	,736

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figuur 5.12 *comprehension questions*

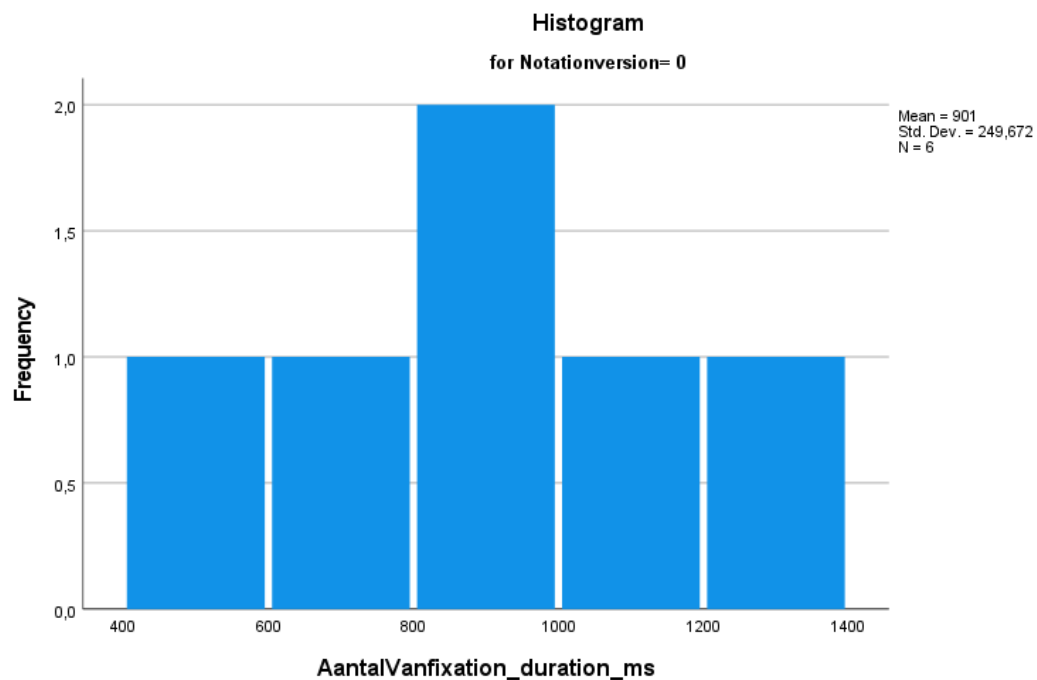
Tests of Normality							
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	versie	Statistic	df	Sig.	Statistic	df	Sig.
fixatieduur	0	,186	8	,200*	,923	8	,457
	1	,212	7	,200*	,912	7	,411
aantalfixaties	0	,166	8	,200*	,936	8	,568
	1	,232	7	,200*	,902	7	,341

*. This is a lower bound of the true significance.

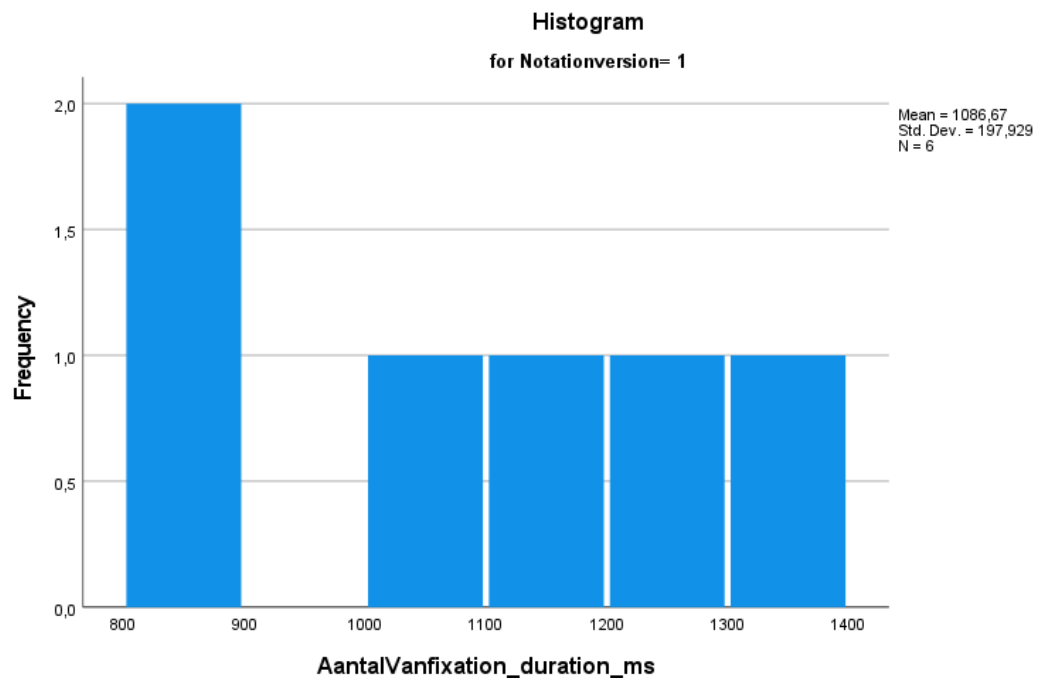
a. Lilliefors Significance Correction

Figuur 5.13 *problem solving tasks*

De test Shapiro-Wilk test geeft aan dat de data een normale verdeling heeft, echter kijken we naar de histogrammen (Figuur 5.14a en b) dan zien we verschillen in distributie. In zo'n geval wordt een niet-parametrische test gebruikt. In ons geval is dat de Mann-Whitney U test.



Figuur 5.14a



Figuur 5.14b

11.3. Sectie *comprehension questions* en *problem solving tasks* data preparatie AOI data

Dezelfde handelingen zijn toegepast in het structureren van de data voor de hypothesen H_a -AOI fixatie duur en H_a -AOI *run count*.

aoi_name	aoi_fixation_total_count	aoi_fixation_average_duration_ms
Fvraag1	2	151
Fvraag1	2	117
Fvraag1	1	104
Fvraag2	1	230
Fvraag2	1	138
Fvraag2	2	184
Gvraag1	1	124
Gvraag1	1	206
VerbindingA-Bvraag2	1	238
VerbindingB-Avraag1	1	166
VerbindingB-Avraag1	1	216
VerbindingB-Avraag2	2	189
VerbindingC-Bvraag1	2	213
VerbindingC-Bvraag1	2	160
VerbindingC-Bvraag1	2	166
VerbindingC-Bvraag1	2	203
VerbindingC-Bvraag1	1	122
VerbindingC-Bvraag1	1	304
VerbindingC-Bvraag2	2	163
VerbindingE-Dvraag1	1	213

Figuur 5.14

aoi_fixation_average_duration_ms bevat de gemiddelde hoeveelheid aantal milliseconden (ms) dat besteed is gedurende een bezoek aan de AOI, waarbij *aoi_fixation_total_count* het totaal aantal keren aangeeft dat een AOI is bezocht. De *aoi_name* is de labelnaam die handmatig is aangemaakt elke keer dat de ogen het AOI gebied binnenkomt.

In figuur 5.14 zien we drie maal Fvraag1 waarbij de eerste twee keer de ogen vijf keer dit gebied zijn binnengegaan. Fvraag1 wordt geaggregeerd waarbij het resultaat als volgt eruit ziet

<i>aoi_name</i>	<i>aoi_fixation_total_count</i>
Fvraag1	5

Merk op dat de dataset alleen de relevante AOI gebieden heeft geregistreerd die bij een vraag horen wanneer het oog op dat moment daar naar kijkt.

In de data set is de *aoi_name* variabele gecodeerd naar categorische data. Bijv. In het bovenstaande voorbeeld heeft dan Fvraag1 de waarde 1 en Fvraag2 de waarde 2, etc..

Controle op normale verdeling

Er is voor deze dataset de Shapiro-Wilk test uitgevoerd om na te gaan of de variabelen een normale distributie volgen. Figuren 5.15a en b tonen de resultaten van de test voor respectievelijk sectie 2 en 3

Tests of Normality							
	Notationversion	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
SomVanaoi_fixation_total_count	Old	,319	6	,056	,785	6	,043
	New	,324	6	,049	,803	6	,063
SomVanaoi_fixation_averageduration_ms	Old	,251	6	,200*	,874	6	,244
	New	,288	6	,132	,859	6	,186

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figuur 5.15a *comprehension questions*

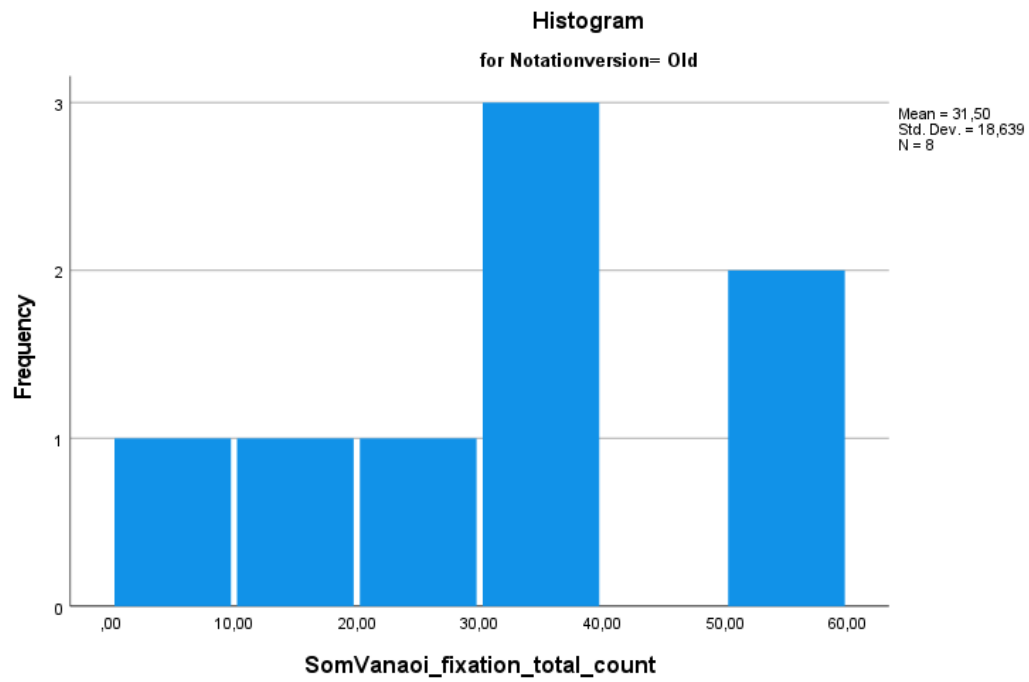
Tests of Normality							
	Notationversion	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
SomVanaoi_fixation_total_count	Old	,176	8	,200*	,943	8	,641
	New	,153	7	,200*	,935	7	,594
SomVanaoi_fixation_averageduration_ms	Old	,145	8	,200*	,947	8	,684
	New	,109	7	,200*	,984	7	,977

*. This is a lower bound of the true significance.

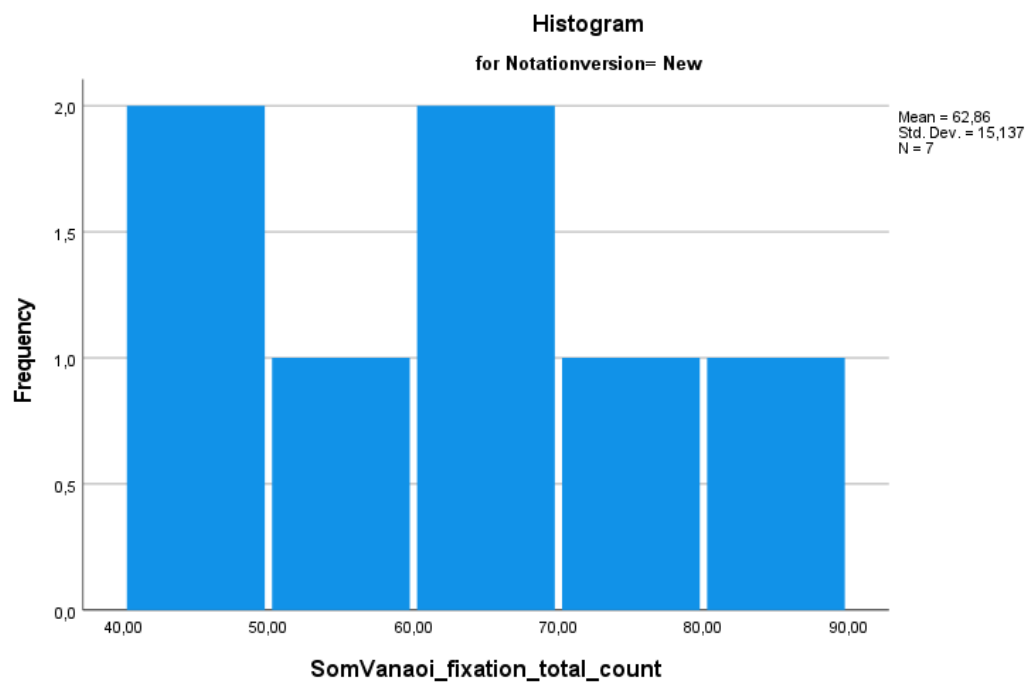
a. Lilliefors Significance Correction

Figuur 5.15b Sectie 3 *problem solving tasks*

De test Shapiro-Wilk test geeft aan dat de data een niet-normale verdeling heeft bij ($p = .043$) (comprehension questions) in de bij de oude notatie en bij de overige data zowel bij de *comprehension questions* als bij de *problem solving tasks* een normale verdeling aangeeft ($p > .05$). Bekijken we een aantal histogrammen dan zie we dat dat beeld niet helemaal klopt (figuur 5.15c).



Figuur 5.15c



Figuur 5.15c



De losse vragen van de comprehension questions en problem solving tasks vertonen allen een niet-normale verdeling daarom worden deze variabelen ook getoetst met de Mann-Whitney U test.

11.4. Sectie *comprehension questions* en *problem solving tasks* data preparatie effectiviteit







Voor de effectiviteit is de survey data van Limesurvey gebruikt. De *Comprehension questions* bestonden uit meerkeuze vragen. De antwoorden zijn gecodeerd met numerieke nominale waarden. Zie figuur 5.17a en b respectievelijk de originele vraag en gecodeerde versie.

 Welk klantdoel wordt tergesteld	 Hoeveel waardeproposities worden in dit model weergegeven
Verlagen financiële druk	5
Verlagen financiële druk	5
Onkosten efficiënter afhandelen	5

Figuur 5.17a

 Vraag1	 Vraag2
1	0
1	0
0	0


Figuur 5.17b

 Vraag1	 Vraag2	 Vraag3	 Vraag4	 Vraag5	 Vraag6	 Vraag7	 Vraag8	 Vraag9	 Vraag10	 Notationw ersion	 TotaalScoreParticipant
1	0	0	1	0	0	0	1	1	0	0	4
1	0	0	0	0	0	0	1	1	0	0	3

Figuur 5.18

Figuur 5.18 laat zien welke bewerking is gemaakt om de effectiviteit te meten. Deze stap is in zowel bij de *problem solving tasks* als bij de *comprehension questions* genomen. Voor iedere groep wordt per respondent de 'goed' scores van elke vraag opgeteld. Dat resultaat komt in de variabele TotaalScoreRespondent terecht.

Voor de *Problem solving tasks* is een andere codering gebruikt want deze bestond uit open vragen hierdoor was de kans aanwezig dat respondenten ook een 'gedeeltelijk goed' konden scoren. Dit kreeg dan de waarde van 0.5. Hier is de volgende codering gebruikt 0 = onjuist, 0.5 = gedeeltelijk juist en 1 = juist.

 Vraag1	 Vraag2	 Vraag3	 TotaalscorePer Participant
0	,0	0	,0
0	1,0	0	1,0
0	,5	0	,5
1	1,0	0	2,0

Figuur 5.20

Controle op normale verdeling

De verdeling van de vraag variabelen zijn getoetst d.m.v. Shapiro-Wilk test waaruit bleek dat de vragen zowel bij de *comprehension questions* als bij de *problem solving tasks* een niet-normale verdeling laten zien evenals de variabele TotaalScoreRespondent.

11.1. Sectie *comprehension questions* en *problem solving tasks* data preparatie efficiëntie

Zoals al eerder vermeld in paragraaf 10.1 is de responstijd in Lymesurvey gecorrigeerd. Dit heeft te maken met het feit dat de meeste respondenten tijdens het maken van de *eye-tracking* test werden geconfronteerd met het her-kalibreren van het systeem wanneer het de ogen niet meer kon volgen.

Controle op normale verdeling

Tests of Normality							
	Notatieversie	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Totaaltijd	0	,180	6	,200 [*]	,953	6	,765
	1	,184	6	,200 [*]	,915	6	,471
Groepstijd: Vraag 1	0	,202	6	,200 [*]	,943	6	,686
	1	,203	6	,200 [*]	,946	6	,704
Groepstijd: Vraag 2	0	.	6	.	.	6	.
	1	,313	6	,068	,832	6	,113
Groepstijd: Vraag 3	0	,198	6	,200 [*]	,945	6	,703
	1	,225	6	,200 [*]	,888	6	,307
Groepstijd: Vraag 4	0	,168	6	,200 [*]	,912	6	,448
	1	,313	6	,068	,780	6	,038
Groepstijd: Vraag 5	0	,202	6	,200 [*]	,973	6	,910
	1	,151	6	,200 [*]	,968	6	,876
Groepstijd: Vraag 6	0	,292	6	,120	,837	6	,123
	1	,245	6	,200 [*]	,919	6	,497
Groepstijd: Vraag 7	0	,188	6	,200 [*]	,900	6	,377
	1	,153	6	,200 [*]	,952	6	,753
Groepstijd: Vraag 8	0	,180	6	,200 [*]	,945	6	,696
	1	,263	6	,200 [*]	,903	6	,393
Groepstijd: Vraag 9	0	,248	6	,200 [*]	,812	6	,075
	1	,172	6	,200 [*]	,923	6	,526
Groepstijd: Vraag 10	0	,200	6	,200 [*]	,918	6	,492
	1	,183	6	,200 [*]	,944	6	,688

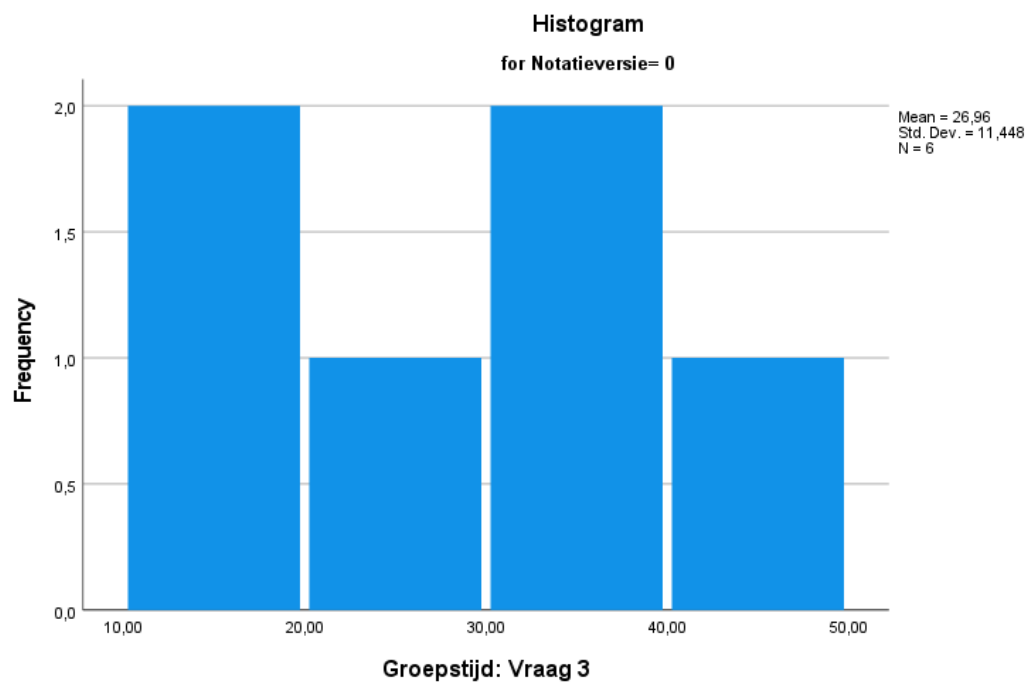
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

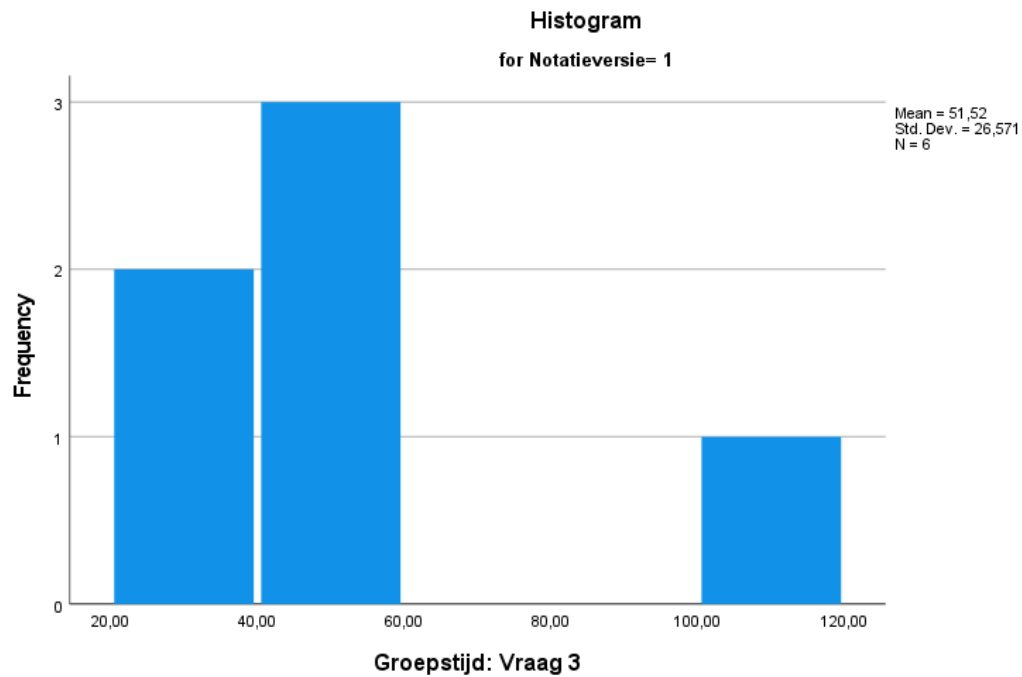
Figuur 5.21

De Shapiro-Wilk test geeft aan dat een groot deel van de variabelen voldoet aan een normale verdeling (zie figuur 5.21). Echter wanneer we kijken naar de visualisatie van een aantal van deze variabelen zien we niet echt een normale verdeling (zie figuur 5.22).

Histograms

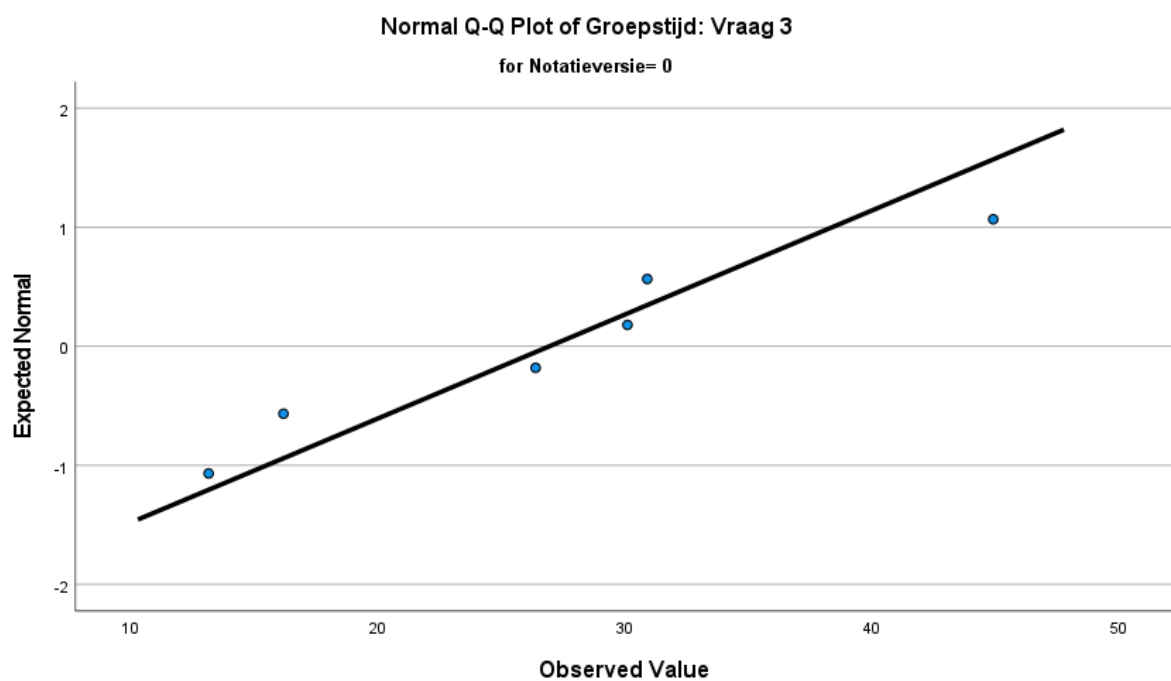


Figuur 5.22a

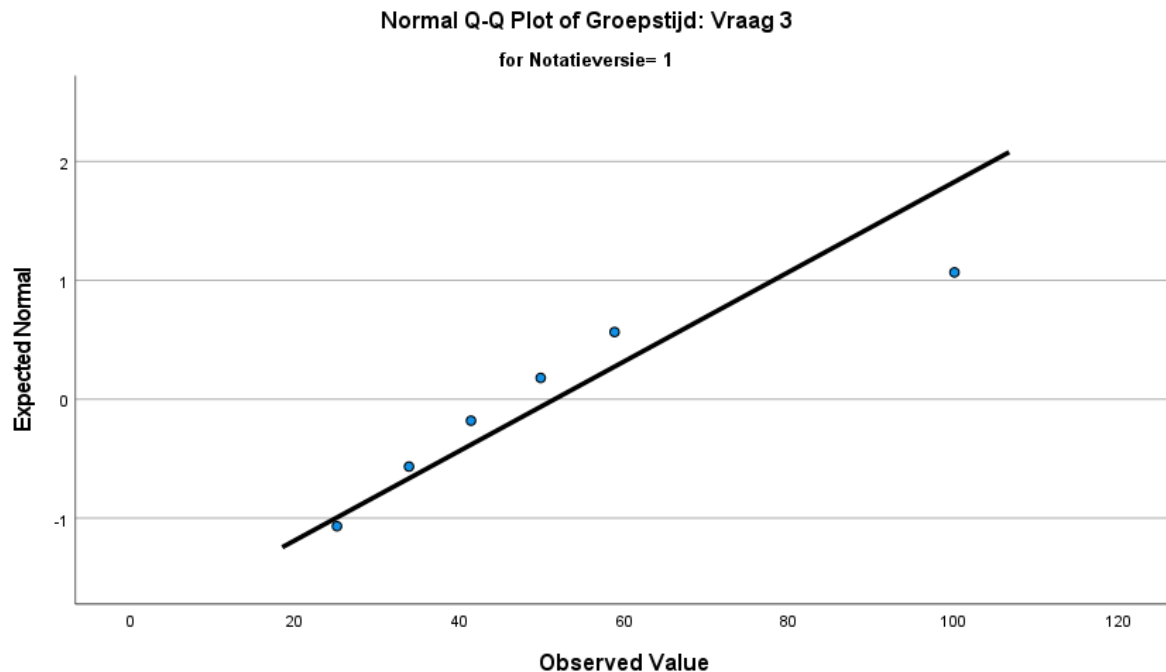


Figuur 5.22b

Zien we niet echt een normale verdeling. De Q-Q plots visualiseren dit nog meer.



Figuur 5.22c



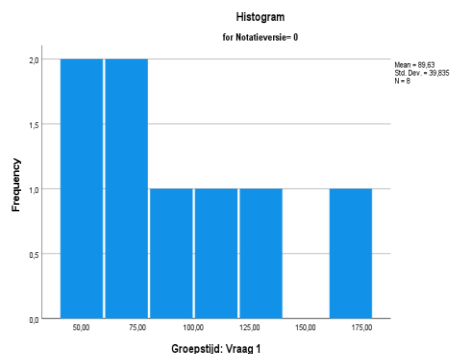
Figuur 5.22d

Een aantal data punten raken de lijn. Vanwege het feit dat de steekproef vrij klein is van omvang en er dus maar een paar data punten zijn, wordt het al snel een normale verdeling als die paar punten voldoen aan een normale verdeling. Daarom wordt in dit geval toch gekozen voor de Mann Whitney U test. Hetzelfde geldt ook voor sectie 3. Figuur 5.23 toont de uitslag van Shapiro-Wilk test:

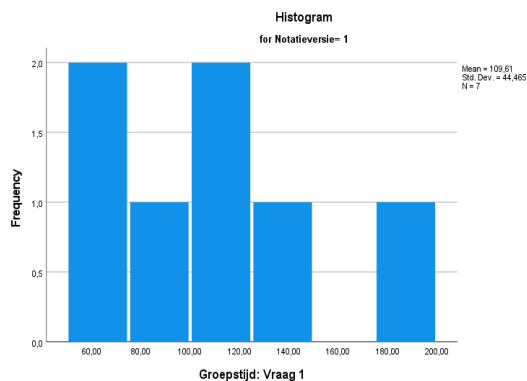
Tests of Normality							
	Notatieversie	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Totaaltijd	0	,170	8	,200 [*]	,960	8	,805
	1	,185	7	,200 [*]	,957	7	,796
Groepstijd: Vraag 1	0	,178	8	,200 [*]	,923	8	,455
	1	,194	7	,200 [*]	,953	7	,759
Groepstijd: Vraag 2	0	,147	8	,200 [*]	,947	8	,678
	1	,243	7	,200 [*]	,855	7	,136
Groepstijd: Vraag 3	0	,150	8	,200 [*]	,921	8	,436
	1	,189	7	,200 [*]	,945	7	,681

Figuur 5.23

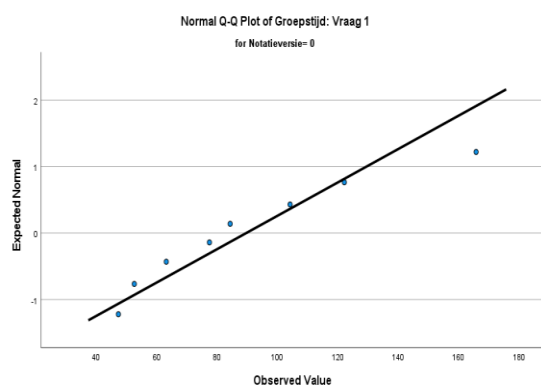
Ook hier geeft de test geen werkelijk beeld van de distributie. Zo laten de histogrammen duidelijk een niet-normale verdeling zien (figuren 5.24a en b). In plaats daarvan wordt er voor deze kleine steekproef grootte de Q-Q plot geraadpleegd om te zien of er een normale verdeling is (zie figuren 5.24c en d).



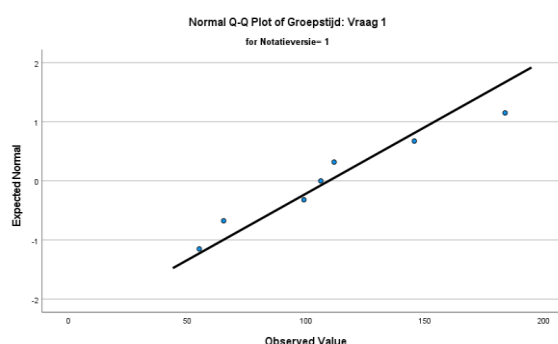
Figuur 5.24a



figuur 5.24b



Figuur 5.24c



Figuur 5.24d

Omdat de data duidelijk niet normaal-verdeeld is wordt de Mann-Whitney test uitgevoerd.

11.2. Sectie comprehension questions en problem solving tasks data preparatie ease of use

Na afloop van het experiment zal aan de hand van een vragenlijst de respondenten gevraagd worden naar het gebruiksgemak van de notatie.

Welken notatie versie heb je eerder gehad	Het was voor mij gemakkelijk te begrijpen wat het model wild emodelleren	Het gebruik van het model was vaak frustrerend	Over het algemeen was het gebruik van het model eenvoudig te gebruiken	Het len
0	1	2	1	
0	4	2	2	
0	0	4	0	
0	2	4	2	
0	1	5	1	
1	1	2	1	
0	4	2	2	
0	5	5	4	
0	2	5	1	
1	4	2	4	
1	1	2	4	

Figuur 5.25

Figuur 5.25 laat de gecodeerde uitkomsten zien van de vragen gesteld in sectie 4. Codering verloopt als volgt:

Vraag over gebruik van de modelleertaal

Ⓛ Dit is een verplichte vraag
Ⓛ Vul alle onderdelen in

	1 - Zeker niet mee eens	2	3	4 - Niet mee eens of oneens	5	6	7 - Zeker mee eens
Het was voor mij gemakkelijk te begrijpen wat het model wilde modelleren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Het gebruik van het model was vaak frustrerend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Over het algemeen was het gebruik van het model eenvoudig te gebruiken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Het leren lezen van het model was makkelijk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

De waarde “1- Zeker niet mee eens” krijgt het getal 0, 2 wordt 1, etc. Behalve bij vraag 2 omdat deze vraag een ontkennende vraag is wordt de codering omgedraaid. Hoe hoger de score hoe beter de nieuwe notatie scoort. Wanneer we vraag 2 niet omdraaien krijgen we een hoog aantal wanneer een respondent aangeeft erg gefrustreerd te raken van het model.

Controle op normale verdeling

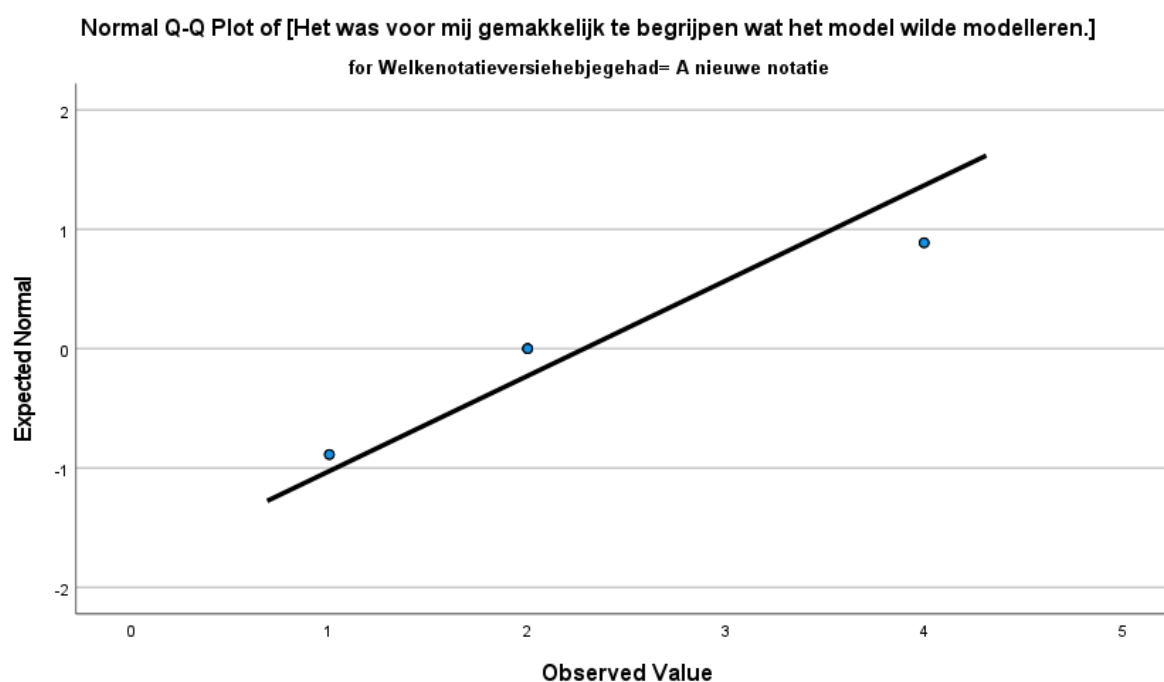
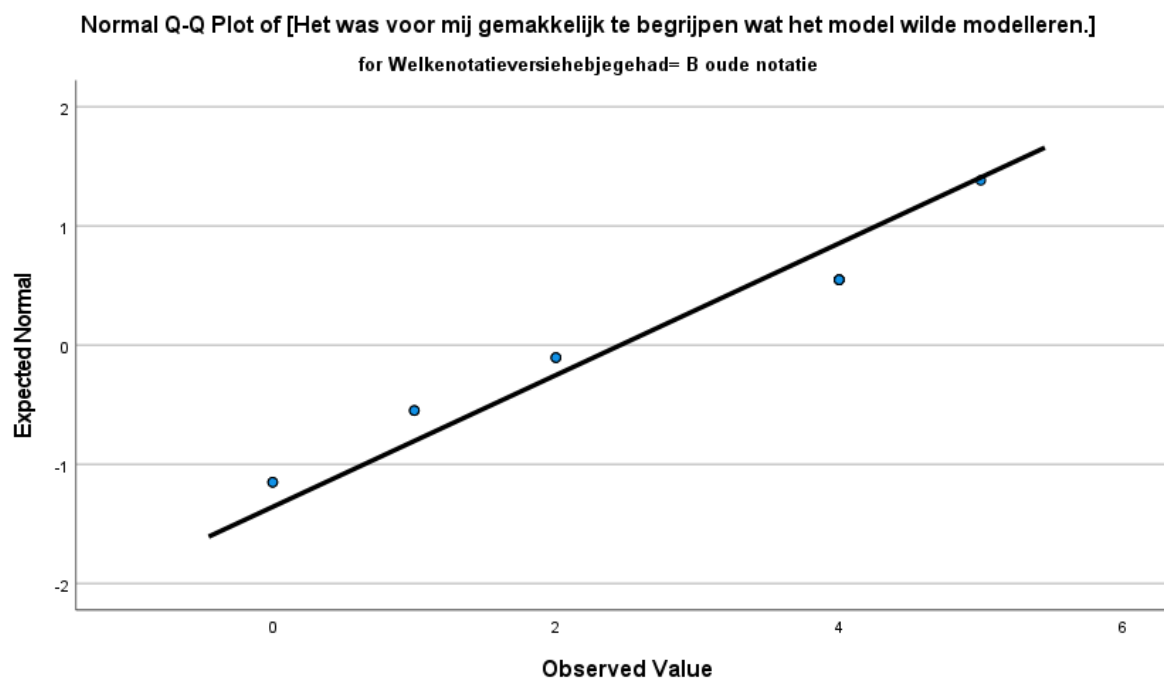
Er is gebruik gemaakt van de Shapiro-Wilk test om te testen of de data normaal verdeeld is.

Tests of Normality							
	Welke notatie versie heb je gehad?	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
[Het was voor mij gemakkelijk te begrijpen wat het model wilde modelleren.]	B oude notatie	,258	11	,039	,882	11	,111
	A nieuwe notatie	,304	7	,049	,819	7	,062
[Het gebruik van het model was vaak frustrerend]	B oude notatie	,269	11	,025	,735	11	,001
	A nieuwe notatie	,447	7	,000	,659	7	,001
[Over het algemeen was het gebruik van het model eenvoudig te gebruiken]	B oude notatie	,208	11	,200*	,875	11	,091
	A nieuwe notatie	,352	7	,009	,760	7	,016
[Het leren lezen van het model was makkelijk]	B oude notatie	,260	11	,036	,863	11	,063
	A nieuwe notatie	,352	7	,009	,760	7	,016

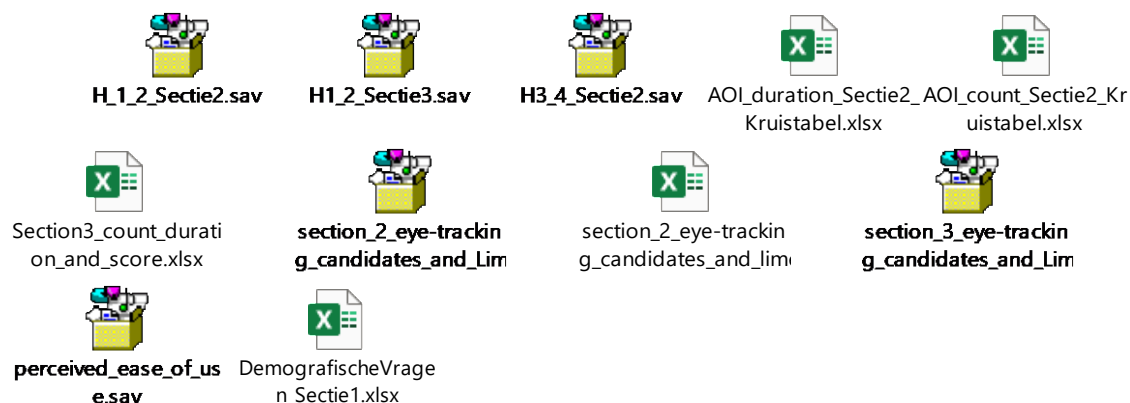
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

De vraag “[Het was voor mij gemakkelijk te begrijpen wat het model wilde modelleren.]” heeft als uitkomst een significantie waarde $p = 0.111$ voor de oude en $p = 0.62$ voor de nieuwe notatie. De Q-Q plots van deze data geven een ander beeld. Dit heeft ons doen besluiten om ook hiervoor de Mann Whitney test te gebruiken.



11.3. Data bestanden



12. Bijlage 6 Effectgrootte berekening

Om te bepalen wat de omvang van de invloed (*effect size*) van de afhankelijke variabele is op de onafhankelijke variabele berekenen we de waarde r (*effect size*). De formule hiervoor is $r = z_{score} / \sqrt{n}$. De uitkomst van deze variabele kan dan vergeleken worden met de *Cohens effect size* tabel.

Effect size	r
Small	0.10
Medium	0.30
Large	0.50

Een effectgrootte van bijv. ≤ 0.30 betekent dat de variabele een middelmatig effectgrootte heeft.

13. Bijlage 7 Communicatie aan respondenten



Groep D: Onderzoek intuïtie



Beste

Hieronder vind je, zoals beloofd, de nodige informatie om deel te nemen aan het onderzoek.

Toegangscode :

Sectie 1
<http://limesurvey.ou.nl/index.php/267398?lang=nl>

Sectie 2
<https://www.realeye.io/test/8d3818ba-4503-436d-8b11-c4f4b1dfc8a2/run>

Sectie 3
<https://www.realeye.io/test/60143dd3-5ce7-4e2b-99e2-320799ed84bb/run>

Sectie 4
<http://limesurvey.ou.nl/index.php/957168?lang=nl>

Mocht je nog vragen hebben dan hoor ik dat uiteraard graag.

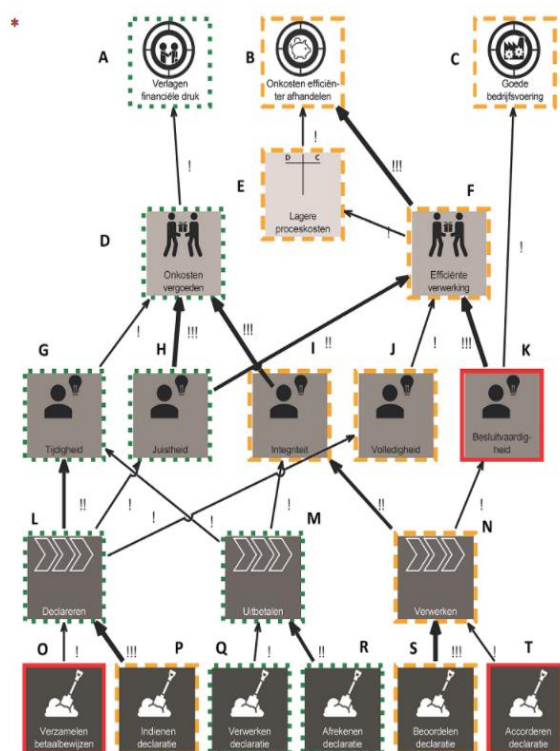
Beantwoorden Allen beantwoorden Doorstu
ma 26

Mail voorbeeld naar studenten

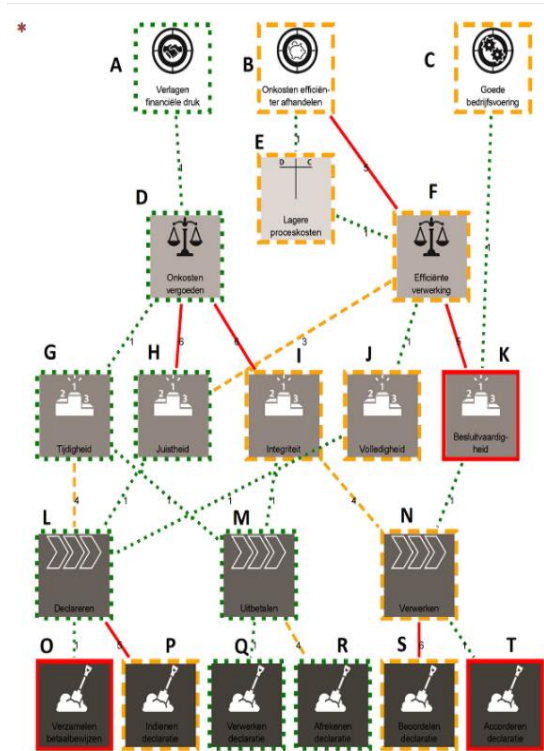
Einde van Sectie 2 Comprehension Questions. Sluit de eye-tracking software door in de browser het tabblad met het record icoontje te kiezen en vervolgens onderaan op de link te klikken. Klik vervolgens, in de mail die je hebt ontvangen, op de volgende link om naar de Problem Solving tasks te gaan.

Eind bericht na het einde van een sectie (in dit geval van Sectie 2). De respondent krijgt instructie.

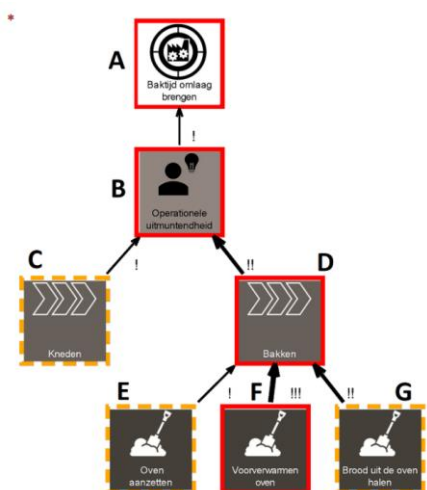
14. Bijlage 8 PGA modellen



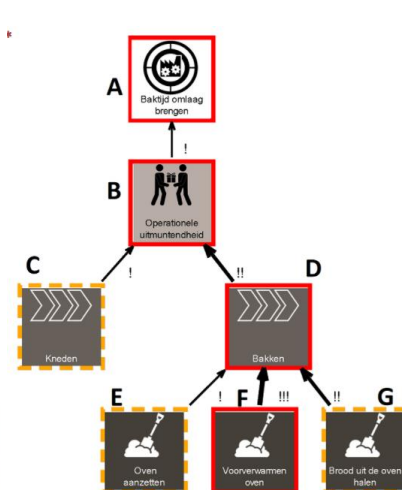
Nieuwe notatie comprehension questions



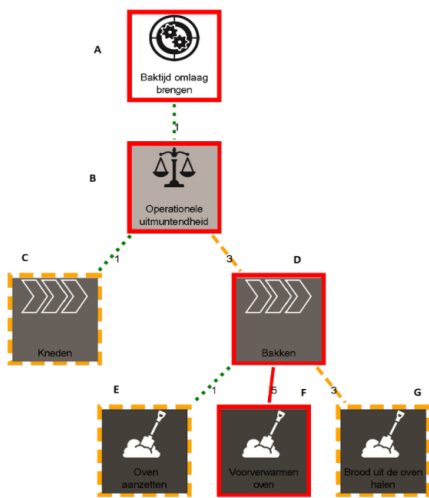
Oude notatie comprehension questions



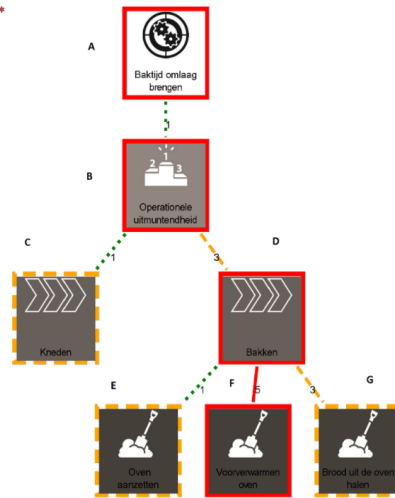
vraag 1 nieuwe notatie Problem solving tasks



vraag 2 nieuwe notatie Problem solving tasks



vraag 1 oude notatie *Problem solving tasks*



vraag 2 oude notatie *Problem solving tasks*



vraag 3 nieuwe notatie *Problem solving tasks*



vraag 3 oude notatie *Problem solving tasks*